Optimized α/β pulse shape discrimination in Borexino

D. Basilico,[1] G. Bellini,[1] J. Benziger,[2] R. Biondi,[3,a] B. Caccianiga,[1] F. Calaprice,[4] A. Caminata,[5] A. Chepurnov,[6] D. D'Angelo,[1] A. Derbin,[7] A. Di Giacinto,[3] V. Di Marcello,[3] X. F. Ding,[4,b] A. Di Ludovico,[4,c] L. Di Noto,[5] I. Drachnev,[7] D. Franco,[8] C. Galbiati,[4,9] C. Ghiano,[3] M. Giammarchi,[1] A. Goretti,[4] M. Gromov,[6] D. Guffanti,[10,d] Aldo Ianni,[3] Andrea Ianni,[4] A. Jany,[11] V. Kobychev,[12] G. Korga,[13,14] S. Kumaran,[15,16,e] M. Laubenstein,[3] E. Litvinovich,[17,18] P. Lombardi,[1] I. Lomskaya,[7] L. Ludhova,[15,16] I. Machulin,[17,18] J. Martyn,[10] E. Meroni,[1] L. Miramonti,[1] M. Misiaszek,[11] V. Muratova,[7] R. Nugmanov,[17] L. Oberauer,[19] V. Orekhov,[10] F. Ortica,[20] M. Pallavicini,[5] L. Pelicci,[15,16] Ö. Penek,[15,f] L. Pietrofaccia,[4,c] N. Pilipenko,[7] A. Pocar,[21] G. Raikov,[17] M. T. Ranalli,[3] G. Ranucci,[1] A. Razeto,[3] A. Re,[1] N. Rossi,[3] S. Schönert,[19] D. Semenov,[7] G. Settanta,[15,g] M. Skorokhvatov,[17,18] A. Singhal,[15,16] O. Smirnov,[22] A. Sotnikov,[22] R. Tartaglia,[3] G. Testera,[5] E. Unzhakov,[7] A. Vishneva,[22] R. B. Vogelaar,[23] F. von Feilitzsch,[19] M. Wojcik,[11] M. Wurm,[10] S. Zavatarelli,[5] K. Zuber,[24] and G. Zuzel[11]

(BOREXINO Collaboration)[*]

[1]*Dipartimento di Fisica, Università degli Studi e INFN, 20133 Milano, Italy*
[2]*Chemical Engineering Department, Princeton University, Princeton, New Jersey 08544, USA*
[3]*INFN Laboratori Nazionali del Gran Sasso, 67010 Assergi (AQ), Italy*
[4]*Physics Department, Princeton University, Princeton, New Jersey 08544, USA*
[5]*Dipartimento di Fisica, Università degli Studi e INFN, 16146 Genova, Italy*
[6]*Lomonosov Moscow State University Skobeltsyn Institute of Nuclear Physics, 119234 Moscow, Russia*
[7]*St. Petersburg Nuclear Physics Institute NRC Kurchatov Institute, 188350 Gatchina, Russia*
[8]*APC, Université de Paris, CNRS, Astroparticule et Cosmologie, Paris F-75013, France*
[9]*Gran Sasso Science Institute, 67100 L'Aquila, Italy*
[10]*Institute of Physics and Excellence Cluster PRISMA+, Johannes Gutenberg-Universität Mainz, 55099 Mainz, Germany*
[11]*M. Smoluchowski Institute of Physics, Jagiellonian University, 30348 Krakow, Poland*
[12]*Institute for Nuclear Research of NASU, 03028 Kyiv, Ukraine*
[13]*Department of Physics, Royal Holloway University of London, Egham, Surrey,TW20 0EX, United Kingdom*
[14]*Institute of Nuclear Research (Atomki), Debrecen, Hungary*
[15]*Institut für Kernphysik, Forschungszentrum Jülich, 52425 Jülich, Germany*
[16]*III. Physikalisches Institut B, RWTH Aachen University, 52062 Aachen, Germany*
[17]*National Research Centre Kurchatov Institute, 123182 Moscow, Russia*
[18]*National Research Nuclear University MEPhI (Moscow Engineering Physics Institute), 115409 Moscow, Russia*
[19]*Physik-Department, Technische Universität München, 85748 Garching, Germany*
[20]*Dipartimento di Chimica, Biologia e Biotecnologie, Università degli Studi e INFN, 06123 Perugia, Italy*
[21]*Amherst Center for Fundamental Interactions and Physics Department, University of Massachusetts, Amherst, Massachusetts 01003, USA*
[22]*Joint Institute for Nuclear Research, 141980 Dubna, Russia*
[23]*Physics Department, Virginia Polytechnic Institute and State University, Blacksburg, Virginia 24061, USA*
[24]*TU Dresden, Dresden, Germany*

[*]spokesperson-borex@lngs.infn.it
[a]Present address: Max-Planck-Institut für Kernphysik, 69117 Heidelberg, Germany.
[b]Present address: IHEP Institute of High Energy Physics, 100049 Beijing, China.
[c]Present address: INFN Laboratori Nazionali del Gran Sasso, 67010 Assergi (AQ), Italy.
[d]Present address: Dipartimento di Fisica, Università degli Studi e INFN Milano-Bicocca, 20126 Milano, Italy.
[e]Present address: Department of Physics and Astronomy, University of California, Irvine, California, USA.
[f]Present address: GSI Helmholtzzentrum für Schwerionenforschung GmbH, 64291 Darmstadt, Germany.
[g]Present address: Istituto Superiore per la Protezione e la Ricerca Ambientale, 00144 Roma, Italy.

(Received 25 January 2024; accepted 9 May 2024; published 14 June 2024)

Borexino could efficiently distinguish between α and β radiation in its liquid scintillator by the characteristic time profile of its scintillation pulse. This α/β discrimination, first demonstrated on the ton scale in the counting test facility prototype, was used throughout the lifetime of the experiment between 2007 and 2021. With this method, the α events are identified and subtracted from the solar neutrino events similar to β. This is particularly important in liquid scintillators, as the α scintillation is strongly quenched. In Borexino, the prominent ^{210}Po decay peak was a background in the energy range of electrons scattered from ^{7}Be solar neutrinos. Optimal α/β discrimination was achieved with a *multilayer perceptron neural network*, with a higher ability to leverage the timing information of the scintillation photons detected by the photomultiplier tubes. An event-by-event, high efficiency, stable, and uniform pulse shape discrimination was essential in characterizing the spatial distribution of background in the detector. This benefited most Borexino measurements, including solar neutrinos in the pp chain and the first direct observation of the CNO cycle in the Sun. This paper presents key milestones in α/β discrimination in Borexino as a term of comparison for current and future large liquid scintillator detectors.

DOI: 10.1103/PhysRevD.109.112014

I. INTRODUCTION

For as long as it operated, Borexino was the only detector capable of measuring solar neutrino interactions (position and energy) on an event-by-event basis with a threshold $\gtrsim$150 keV, i.e., down to the ^{14}C β-spectrum endpoint. An important feature of Borexino was the possibility of efficiently separate events initiated by recoiling electrons (β-like events) versus α particles. The former include solar neutrino interactions, as well as background from β and γ decays. This was possible via pulse-shape discrimination (PSD) techniques that exploit the different time profile of scintillation emission for α and β-like events (see, e.g., [1]). The so-called α/β discrimination played an important role in solar neutrino measurements throughout the Borexino data taking between 2007 and 2021. It is worth noting that Borexino also achieved β^-/β^+ separation via PSD, as reported in [2,3]; the latter topic is beyond the scope of this article.

PSD for α/β separation was first studied within the Borexino program with the 4-ton "Counting Test Facility" (CTF) prototype [4,5]. The original method is based on the *Gatti* parameter [6] and enabled a statistical subtraction of α background, especially from ^{210}Po, from the measured energy spectrum. Monochromatic, 5.3 MeV ^{210}Po *alphas* ($Q_\alpha = 5407$ keV) appeared in the Borexino liquid scintillator as a peak at ~500 keV of electron-equivalent energy due to a greater than ten-fold quenching of the scintillation for these highly ionizing tracks [5]. At the beginning of the Borexino data taking, the ^{210}Po rate was ~8000 counts per day per 100 tons (hereafter cpd/100 t). Quenching was also observed for other α particles. These include those from the thoron (^{220}Rn) and radon (^{222}Rn) decay chains which are handily identified using their time coincidence, e.g., ^{212}Po (8954 keV), ^{214}Po (7833 keV), and ^{218}Po (6114 keV).

Due to its quenching, the ^{210}Po peak falls within the ^{7}Be solar neutrino Compton-like energy spectrum, which presents a characteristic shoulder at 662 keV. Although in this case the ^{7}Be shoulder appears at higher energy than the ^{210}Po peak, making it possible for the multiparameter spectral fit to clearly identify these two separate components, the Borexino analysis was performed both with and without bin-by-bin statistical α/β subtraction of the ^{210}Po peak to ensure that there was no subtle bias due to the presence of the α background. This was particularly true in phase I of the experiment, when the ^{210}Po activity was more than two orders of magnitude greater than the ^{7}Be event rate (~50 cpd/100 t over the entire energy range). In each bin, we assumed that the Gatti parameter is nearly normally distributed with a mean value that is linearly dependent on energy [7–9].

As data-taking progressed, the ^{210}Po naturally decayed with a lifetime $\tau = 199.6$ d. However, the reduction of this background was counterbalanced by a progressive degradation of the detector energy resolution as a result of the loss of photomultiplier tubes (PMTs). The count of working PMTs decreased from ~2000 units in mid 2007 to ~1000 units at the end of 2021. This effect and the need for a more uniform, stable, and higher efficiency α/β discrimination for the study of CNO solar neutrinos suggested exploring novel techniques based on neural networks already extensively employed in particle physics. We pursued neural networks based on *multilayer perceptron* (MLP). The subject of this paper is the description of the

MLP input parameters, structure, and training strategy given the properties and layout of the Borexino scintillator. It also presents studies of the efficiency of the network for α/β discrimination.

In Sec. II, the main characteristics of the Borexino detector and its main physics results relevant to this article are briefly reviewed. In Sec. III, the α/β PSD in Borexino using the Gatti parameter is presented. In Sec. IV, the implementation strategy of the MLP on the Borexino scintillator time profile is described along with an evaluation of its performance and efficiency. Finally, in Sec. V, the impact of α/β MLP discrimination on the CNO solar neutrino analysis and on other Borexino results over its 14-year lifetime are discussed.

II. THE BOREXINO EXPERIMENT

Borexino was located in the Hall C of Laboratori Nazionali Gran Sasso (LNGS) of the Italian Institute of Nuclear Physics (INFN) [10]. The detector had been taking data from mid-2007 to the end of 2021, and is currently under decommissioning. The detector is made of spherical concentric layers of increasing radiopurity (see [11] for details): the innermost core, called the inner vessel (IV), consists of about 280 tons of liquid scintillator (*pseudocumene* mixed with 1.5 g/l of PPO as scintillating solute) contained inside an ultrapure nylon vessel with a thickness of 125 μm and a radius of 4.25 m. A Stainless Steel Sphere (SSS), filled up with the remaining 1000 m^3 of buffer liquid (*pseudocumene* mixed with *dimethylphthalate* quencher) is instrumented with more than 2000 PMTs to detect scintillation light inside the IV. Ultimately, the SSS is placed inside a water tank (WT) with a volume of approximately 2000 m^3, functioning as a Cerenkov veto system, equipped with 200 PMTs. Using results from the study of internal residual contaminations and from the 2010 calibration, it was found that the detector is capable of determining the event position with an accuracy of ∼10 cm (at 1 MeV) and the event energy with a resolution following approximately the relation $\sigma(E)/E \simeq 5\%/\sqrt{E/[\mathrm{MeV}]}$.

The Borexino dataset is divided into three different phases: phase I, from May 2007 to May 2010, ended with the calibration campaign, in which the first measurement of the ^{7}Be solar neutrino interaction rate [7–9] and the first evidence of the *pep* solar neutrinos [2] were performed; phase II, from December 2011 to May 2016, started after an intense purification campaign with unprecedented reduction of the scintillator radioactive contaminants, in which the first spectroscopic observation of the *pp* neutrinos with 10% precision was published [12], and was later updated in the comprehensive analysis of all *pp* chain neutrino fluxes [3,13,14]; finally, phase III, from July 2016 to October 2021, started after the thermal stabilization program, in which the first detection of the CNO neutrinos [15] and its subsequent improvements [16,17] were achieved.

TABLE I. Solar neutrino interaction rates in Borexino and extrapolated solar neutrino fluxes for the different components of the *pp* chain and CNO cycle. Rates are reported in cpd/100 t, while fluxes are reported in cm^{-2} s^{-1}. N.B.: HZ stands for *high metallicity* assumption.

Species	Rate [cpd/100 t]	Flux [cm^{-2} s^{-1}]
pp	$(134 \pm 10)^{+6}_{-10}$	$(6.1 \pm 0.5)^{+0.3}_{-0.5} \times 10^{10}$
^{7}Be	$(48.3 \pm 1.1)^{+0.4}_{-0.7}$	$(4.99 \pm 0.11)^{+0.06}_{-0.08} \times 10^9$
pep (HZ)	$(2.7 \pm 0.4)^{+0.1}_{-0.2}$	$(1.3 \pm 0.3)^{+0.1}_{0.1} \times 10^8$
^{8}B (> 3 MeV)	$0.223^{+0.021}_{-0.022}$	$5.68^{+0.42}_{-0.44} \times 10^6$
hep	<0.002 (90% CL)	$<1.8 \times 10^5$ (90% CL)
CNO	$6.7^{+1.2}_{-0.8}$	$6.7^{+1.2}_{-0.8} \times 10^8$

The most important solar neutrino results in terms of interaction rate and corresponding fluxes are summarized in I. Thanks to its unprecedented radio-purity, Borexino has also set a lot of limits on rare processes, such as potential electron decay [18], non-standard neutrino interaction [19], high energy neutrinos correlated with astrophysical events [20], neutrino magnetic moment [21] and sterile neutrino [22]; and performed other neutrino physics studies, such as, e.g., geo-neutrino detection (for review, see [23]). As will be highlighted in Sec. V, all these important results are strongly dependent on the α/β PSD optimization. In the following section, the history of α/β discrimination in Borexino is discussed, starting from the first method based on the Gatti parameter.

III. α/β DISCRIMINATION: THE *GATTI* PARAMETER

The α/β discrimination in Borexino is possible thanks to the sizeable difference between the time distributions of the scintillation light (pulse shape) for α and β-like events (see Fig. 1). For each event that meets the threshold condition of a few tens of photoelectrons (PE) in 100 ns (∼60 keV), the arrival times of PEs on each PMTs are recorded. As described in detail in [1], for each detected PE, the arrival time and the charge are measured by an analogue and digital electronics chain. When a trigger occurs, the time and the charge of each PMT that has detected at least one photoelectron in a time gate of 7.2 μs, is recorded. The time is measured by a time-to-digital converter (TDC) with a resolution of about 0.5 ns, while the charge (after integration and pulse shaping) is measured by an 8 bit analog-to-digital converter (ADC). The time resolution is smaller than the intrinsic PMT time jitter of about 1.1 ns.

For each event, time, charge, and position are reconstructed by the offline software. The code identifies a group of time-correlated hits in the recorded time window of 16 μs, called the"cluster." Each event is generally made up of a single cluster, but in the case of fast coincidences like ^{214}Bi-^{214}Po and ^{85}Kr-^{85m}Rb close decays, or in the case of

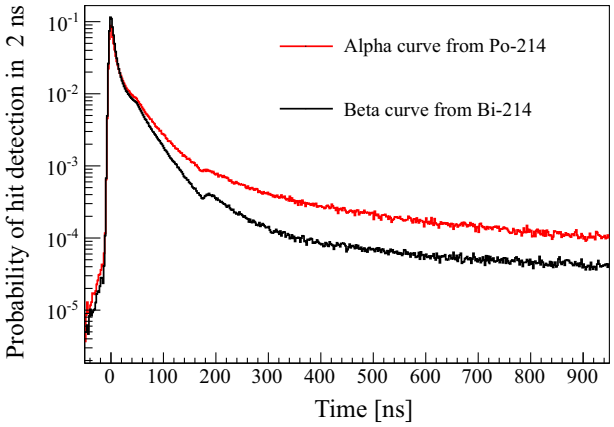

FIG. 1. The reference $P_\alpha(t)$ (red) and $P_\beta(t)$ (black) pulse shapes obtained by tagging the ^{222}Rn-correlated ^{214}Bi-^{214}Po coincidences. The dip observed at 180 ns is the result of the dead time inherent in each electronic channel after the detection of a hit. The small knee around 60 ns is due to the reflected light on the surface of the SSS and on the PMTs photo-cathodes.

accidental pile-ups (often due to the highly frequent ^{14}C events), more than one cluster could be identified. The position of events is determined via a photon time-of-flight maximum likelihood method with probability density functions (PDF) based on experimental data and Monte Carlo simulations, resulting in an uncertainty of 10 cm for each of the three Cartesian spatial coordinates. The spatial resolution is expected to scale naively as $1/\sqrt{\mathrm{NPE}}$ where NPE is the number of detected photoelectrons [24]. For the most important analyses in Borexino, the fundamental event selection is based on the following criteria: internal only trigger (no muon veto coincidence), event time 2 ms away from a preceding muon event, a single cluster in the acquisition window, and the position reconstructed in $r \lesssim 3$ m, where r is the distance from the detector center. These cuts guarantee that the selected event is a neutrino-like candidate, i.e., an event occurred in the innermost part of the IV ($\lesssim$100 tons around the center) and far enough from the external background coming from the SSS and from the IV structures.

After applying the selection criteria listed above, the typical Borexino spectrum shows a prominent ^{210}Po α peak at about 500 keV, which falls within the ^{7}Be energy window; see e.g. [8]. At the beginning of phase I, the ^{210}Po activity was of the order of 10^4 cpd/100 t. At the beginning of phase II, more than 4 years later, the activity decreased by one order of magnitude to $\sim 10^3$ cpd/100 t, a bit more than expected because the water extraction campaign reintroduced a small amount of ^{210}Po. Finally, in phase III, after more than 4 years from the water extraction, the scintillator convective motions were drastically reduced thanks to the thermal insulation campaign (see Sec. V for further details). This significantly reduced ^{210}Po activity by another order of magnitude, reaching $\sim 10^2$ cpd/100 t. This allowed Borexino to reach the ^{210}Po condition required for the CNO measurement [15] via the ^{210}Bi independent constraint from the secular equilibrium, in jargon ^{210}Bi-^{210}Po *link* (see for details Sec. V A).

An estimation of the ^{210}Po activity and its possible independent quantification for the Borexino analysis can be done, for example, by defining a simple parameter called `tail-to-tot` (`t2t`), which is defined as the fractional portion of the time distribution of hits above a given characteristic time t_0 with respect to the beginning of the scintillation, namely,

$$\mathtt{t2t} = \frac{\int_{t_0}^{\infty} S(t)\mathrm{d}t}{\int_{0}^{\infty} S(t)\mathrm{d}t}, \tag{1}$$

where $S(t)$ is the scintillation time distribution. This ideal formula is computed numerically, typically by summing over time bins within a finite time window lasting a few microseconds. Beyond this window, the scintillation is essentially negligible. The characteristic time t_0 can be optimized by maximizing the figure of merit defined as the difference between the means of the `t2t` populations for α and β events. This sort of parameter works very well for example for separating electron and nuclear recoils in liquid argon scintillation chambers [25], where the scintillation light is basically made up of a combination of two typical exponential decay times, differing by 3 orders of magnitude from each other (typically 6 and 1600 ns). This is not the case with the Borexino scintillator, where the time behavior is more complicated and less specific for different types of particles [26]. As a consequence, `t2t` in Borexino gives a milder α/β separation than a real high efficiency event classification.

Instead, a more efficient identification of α/β, can be performed using discriminating procedures such as the Gatti optimal filter [6]. The latter allows one to classify two types of events with different, but known time distributions of hits as a function of time. Their reference shapes $P_\alpha(t)$ and $P_\beta(t)$ are created by averaging the time distributions of a large sample of events selected independently, without any use of pulse shape variables. A practical way to build reference shapes is to use the ^{214}Bi-^{214}Po fast coincidence, originating from the ^{222}Rn events in the scintillator. The ^{214}Bi-^{214}Po coincidences (a few hundreds of μs) in Borexino are tagged by a space-time correlation with about 90% efficiency, basically limited by the trigger threshold of the β event. All the first events within the coincidence window provide a pure β-like sample with events mostly located in the energy interval 1500–3000 keV as superposition of the different β and γ lines, while the second events provide a pure α sample with events peaked at about 800 keV and smeared only by the detector resolution. The radon events in the IV (with about one week lifetime) are strictly related to invasive operations on the scintillator (especially at the beginning of phase I and during the WE

campaign), and are basically absent in quiet periods such as phase II and especially phase III.

The functions $P_\alpha(t)$ and $P_\beta(t)$ represent the PDFs as a function of the time of detection of a PE for events of type α or β, respectively. Let $e(t)$ be the normalized time distribution of the light for each event. The Gatti parameter G is defined as

$$G = \int e(t)w(t)dt, \tag{2}$$

where $w(t)$ are the weights given by

$$w(t) = \frac{P_\alpha(t) - P_\beta(t)}{P_\alpha(t) + P_\beta(t)}. \tag{3}$$

The parameter G follows a probability distribution with a mean value $\langle G_{\alpha,\beta}\rangle$, which depends on the type of particles, that is,

$$\langle G_{\alpha,\beta}\rangle = \int P_{\alpha,\beta}(t)w(t)dt. \tag{4}$$

In the Borexino scintillator, the Gatti mean values are empirically found to be linearly decreasing with energy. Finally, considering the Poissonian statistical fluctuations of the entries in each time bin, the corresponding variance, following the variance expansion identity, reads

$$\sigma^2_{G_{\alpha,\beta}} = \int P^2_{\alpha,\beta}(t)w(t)dt - \langle G_{\alpha,\beta}\rangle^2. \tag{5}$$

In the real experimental case, the integration in Eqs. (4) and (5) is converted into a sum over histograms binned at 1 ns from zero to about 1.5 μs. In the Borexino scintillator, the α pulses are slower and therefore have a longer tail than the β pulses. This basically represents the key for the α/β separation. Examples of reference shapes $P_\alpha(t)$ and $P_\beta(t)$ from the ^{214}Bi-^{214}Po tagging of the radon events from phase I are shown in Fig. 1: the dip at 180 ns is due to the dead time applied to every individual electronic channel after each detected hit. The small knee around 60 ns is due to the reflected light on the SSS surface and on the PMTs photocathodes. The distributions of the corresponding G parameters (G_α and G_β) for events with respect to these reference PDFs are shown in Fig. 2. The two distributions, which are approximately Gaussian-shaped, partially overlap due to the sizeable G variance. As a consequence, when the number of α events largely exceeds that of the β's, a high efficiency event-by-event α/β selection is anyway limited. In principle, a bin-by-bin statistical separation of the two event populations is possible whenever the $G_{\alpha,\beta}$ distributions are known either analytically or through a Monte Carlo simulation. Since the mean values and variances of $G_{\alpha,\beta}$ are energy dependent, their distributions

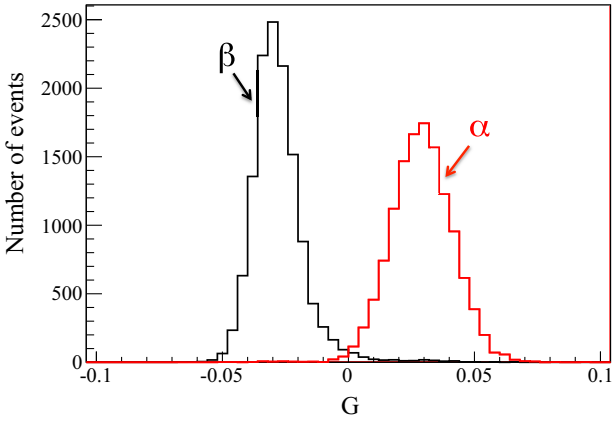

FIG. 2. The distribution of G_α (red) and G_β (black) [see Eq. (2)] for events obtained by tagging the radon correlated ^{214}Bi-^{214}Po coincidences.

are fitted by two Gaussian models for each bin in the energy spectrum of interest, and their values are forcibly constrained around the linear dependence guess. The integrals of the fitted curves represent the relative contribution of each species in each energy bin, and the α contribution is subtracted from the total bin content to obtain the β-like spectrum by statistical subtraction.

One makes a reasonable hypothesis that the underlying distributions are Gaussian. The fit procedure also provides the error of the estimated particle population to be replaced in the corresponding bin in which the subtraction is performed. In bins where one species greatly outnumbers the other, for example in the energy bins containing the ^{210}Po peak, the mean values of the Gaussian parameters are fixed to their predicted values extrapolated from the energy dependency trend in order to avoid any possible bias in the subtraction procedure. Figure 3 shows an example of the

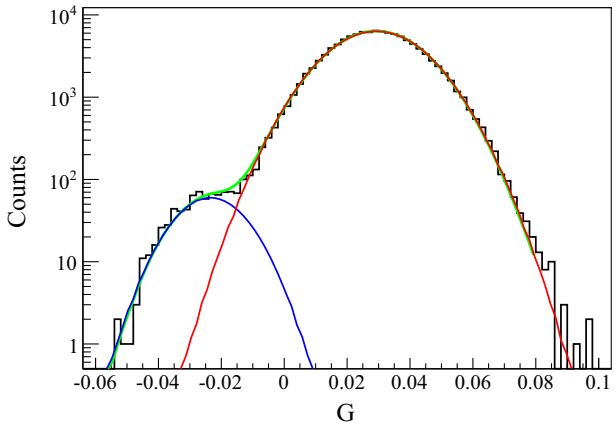

FIG. 3. Example of α/β statistical subtraction with the analytical curves for events in the energy range 200-205 NPE. The blue and red lines show the individual Gaussian fits to the Gatti parameter distributions for the β and α components, respectively, while the green line is the total fit.

$G_{\alpha\beta}$ parameter in the energy range 200-205 NPE and its fit with the analytical model. Furthermore, any other possible double Gaussian fit bias due to the large difference in the two population statistics is corrected using toy Monte Carlo simulations with the same population ratio.

The statistical subtraction can be applied in the entire ^{7}Be energy window, removing all the α background coming mostly from ^{210}Po and other ^{222}Rn α's daughters such as ^{214}Po and ^{218}Po escaping the fast coincidence cut. This secondary subdominant contamination is partially affecting phase I, but is completely negligible in phase II and phase III, thanks to the better background control achieved after the WE campaign and ^{210}Po decay. The error associated with the statistical subtraction is propagated as a systematic uncertainty in the final neutrino interaction rates [7–9]. It is worth mentioning that a possible bias due to the presence of the ^{210}Po peak is not negligible only in phase I when the ^{210}Po activity is much larger than ^{7}Be neutrino interaction rates. In phase II and in phase III the statistical subtraction of the α component is not applied and ^{210}Po is simply quantified by the spectral fit; see [3,15].

Figure 4 shows the Gatti distribution as a function of the event energy in NPE for the first 300 days of Borexino phase II. The big blob on the top represents the α distribution, which basically consists of ^{210}Po events, while the bottom horizontal belt represents the β-like component (solar neutrinos and background). The Gatti parameter shows a neat separation of the α/β population as a function of the energy.

Figure 5 shows the implementation of the α/β statistical subtraction in phase I: the black curve represents the energy distribution of all events before applying the basic selection criteria. The blue curve represents the event energy distribution after fiducial volume selection: below 100 NPE the spectrum is dominated by ^{14}C decay (β^-, $Q = 156$ keV) [27] and the peak at 200 NPE is dominated by ^{210}Po decays. The red curve is the final spectrum after statistical subtraction of the α component. The Compton-like edge around 300 PE is due to ^{7}Be solar neutrinos. Finally, the large peak around ~600 NPE is the spectrum of the cosmogenic ^{11}C (β^+, $Q = 1.98$ MeV, created *in situ* by cosmic ray-induced showers).

Thanks to its good discrimination power, the α/β based on the Gatti parameter has been applied in many Borexino analyses as a *soft cut* for pre-selecting events as well as *hard cut* to locate the main α contaminants and to understand the nature of the main backgrounds [1], for example in the geo-neutrino analysis [23]. In these cases, no statistical subtraction is applied, and the Gatti parameter selects the α population with a given efficiency, depending on the position of the cut.

The optimization of the Gatti filter, already exploited in the Borexino CTF, played a crucial role in many important Borexino studies. Nevertheless, new requirements and some drawbacks pushed the collaboration to investigate other novel techniques based on neural networks. As will be described in Sec. V, the CNO feasibility study required a deep understanding of the spatial evolution of the ^{210}Po contamination since the beginning of phase II. This analysis required a high efficiency event-by-event selection uniform in space, instead of a statistical subtraction that can be easily modeled in energy and time. The PMT loss and the consequent resolution degradation affected the Gatti parameter distributions, but more importantly, the Gatti has always shown a spatial dependence, especially along the radial direction. Figure 6 shows the shift of the $G_{\alpha,\beta}$ as a function of r^3 for ^{214}Bi-^{214}Po events (N.B.: plotting data as a function of r^3 removes the spherical volume dependence over r). This dependence can be neither easily modeled nor fully reproduced in Monte Carlo simulations. In contrast to the Gatti filter, artificial neural networks, which process a multitude of input parameters and output their corresponding rankings for the specific case of α/β selection, have significantly contributed to elucidating the origin of the

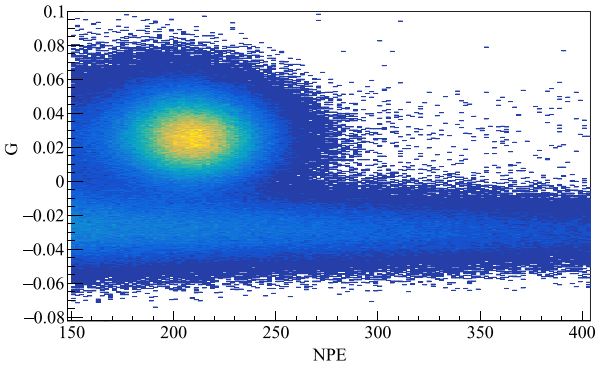

FIG. 4. Example of α/β separation in the Gatti-Energy (NPE) space (first 300 days of Borexino phase II). The big blob on the top represents the α distribution (^{210}Po), while the bottom horizontal belt represents the β-like component.

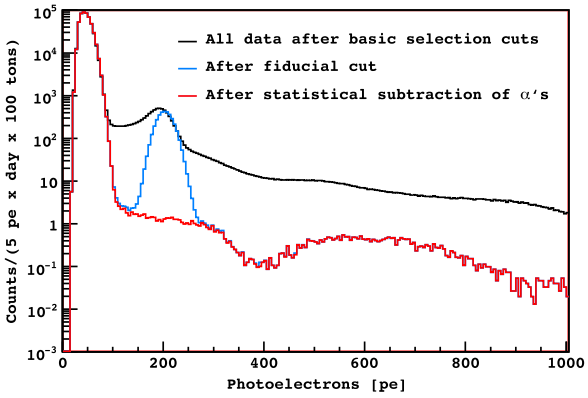

FIG. 5. The raw NPE charge spectrum after the basic selection criteria (black), after the fiducial volume cut (blue), and after the statistical subtraction of the α-emitting contaminants (red).

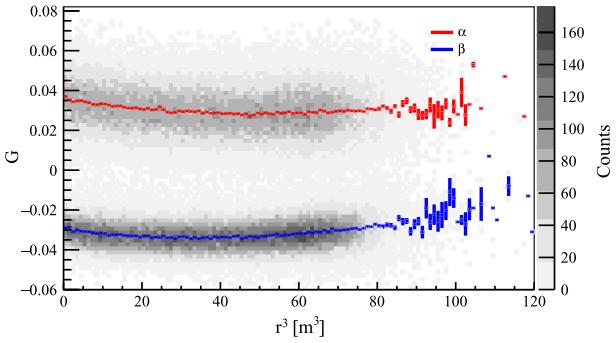

FIG. 6. Radial (r^3) dependence of the Gatti parameter G (gray) for a sample of ^{214}Bi-^{214}Po events. The red and blue points represent the mean values with their uncertainty for α and β-like events, respectively.

radial dependence of the PSD. They facilitate a more uniform selection process, enabling a controllable efficiency dependence on energy and time, as will be elaborated in Sec. IV C. The strategy to implement and tune a class of *multilayer perceptrons* is examined in Sec. IV.

IV. IMPROVING THE SELECTION WITH MLP

A. Artificial neural networks

Artificial neural networks (ANNs) play a crucial role in the field of machine learning as they are specifically engineered to identify patterns within data [28]. This characteristic renders ANNs highly effective for tasks such as classification (organizing data into predefined groups), clustering (identifying common traits in data and grouping them accordingly), and generating predictions from the available data.

An artificial neural network (ANN) is essentially a simulated network of interconnected neurons, where each neuron generates a specific response based on a particular set of input signals. These input signals can represent the attributes of external data samples, like images or documents, or they can originate from other neurons.

By applying an external signal to some input neurons, the network is put into a defined state that can be measured from the response of one or several output neurons. Therefore, one can view the neural network as a mapping from a space of input variables $x_1, \ldots, x_{n_{\text{var}}}$ to a *one-dimensional* (e.g. in the case of a signal-versus-background discrimination problem) or *multidimensional* space of output variables. The mapping is nonlinear if at least one neuron has a nonlinear response to its input. It is crucial to emphasize that the linearity of the Gatti parameter with respect to the input parameters can be readily observed from Eq. (2). Consequently, based on the input dataset, it is unable to surpass its inherent statistical capability.

A multilayer perceptron (MLP) belongs to the category of feed-forward artificial neural networks. The study at hand involves working with an MLP that comprises four layers of nodes: an input layer, two hidden layers, and an output layer. With the exception of the input nodes, each node functions as a neuron employing a non-linear activation function. The training of MLP is carried out using a supervised learning approach known as *back-propagation*. The decision to use this approach was influenced by the straightforward nature of the alpha–beta separation issue, which primarily involves two clusters in the input parameter domain. Adhering to conventional methods, introducing excessive hidden layers would likely result in issues such as overfitting. The evaluation of the chosen network's performance was determined by this decision. However, as will be further explained, the selection of input parameters holds more importance and required a thorough analysis.

B. TMVA package

The toolkit for multivariate analysis (TMVA) provides a ROOT-integrated environment for the processing, parallel evaluation, and application of multivariate classification techniques [29]. TMVA is specifically designed for the needs of high-energy physics (HEP) applications where the search for ever smaller signals in ever larger datasets has become essential to extract a maximum of the available information from the data. Multivariate classification methods based on machine learning techniques have become an essential ingredient in most of the HEP analyses. The package hosts a large variety of multivariate classification algorithm, e.g. artificial neural networks (three different MLPs implementations), support vector machines (SVM), boosted decision trees (BDT), etc.

Before implementing the algorithm, it is crucial to establish separate sets of input data for *training* and *testing* in multivariate techniques. An essential aspect is to pinpoint the key input parameters that are crucial for maximizing the efficiency of signal pulse shape discrimination. This can be accomplished by examining the ranking of the variables in each controlled trial.

C. Selection of input variable and different versions

In the case of α/β discrimination in Borexino, a set of `t2t` variables were defined for ten different t_0, according to Eq. (1). Due to the fact that the distributions for α and β (Fig. 1) differ mainly in the tails, a list of t_0's after 10 ns were optimized as

$$t_0 \in \{35, 70, 105, 140, 175, 210, 245, 280, 315, 350\}\ (\text{ns}). \tag{6}$$

To this set, the root mean square time $\langle\sigma\rangle_t$ (or shortly `RMS`) and `kurtosis` $\langle\kappa\rangle_t$ of the photoelectron time distribution were added, defined respectively as

$$\langle\sigma\rangle_t^2 = \int_0^\infty \hat{S}(t)(t-\langle t\rangle)^2 \mathrm{d}t \tag{7}$$

$$\langle\kappa\rangle_t = \frac{1}{\langle\sigma\rangle_t^4}\int_0^\infty \hat{S}(t)(t-\langle t\rangle)^4 \mathrm{d}t, \tag{8}$$

where $\hat{S}(t)$ is the normalized scintillation PDF and $\langle t\rangle$ is the *mean-time* defined as

$$\langle t\rangle = \int_0^\infty t\hat{S}(t)\mathrm{d}t. \tag{9}$$

At this stage, having a set made of `t2t`'s, `RMS` and `kurtosis`, totalling 12 input parameters, the MLP algorithm returns the discrimination efficiency similar to Gatti, as expected. The Gatti filter has been shown to be the most effective linear filter for two specific reference probability density functions. Therefore, it was essential to explore incorporating extra inputs that were not present in the reconstructed scintillation pattern.

It is important to note that in Borexino there are two ways of defining the scintillation PDF: the first is defined with respect to an absolute reference time (that is, the distribution of the times at which the PMTs are fired); the second involves sorting the same list of times after the position of the event is reconstructed in the detector, facilitated by the measurement of the photon time-of-flights from the point-like events. Indeed, from subsequent trials, it was observed that the mean-time variable of the hits calculated before the position reconstruction (`mean-time0`) was adding some missing information, possibly lost with the position correction (13 input parameters in this case). This recovered information improved the α/β discrimination, even solving the radial dependence observed in the Gatti parameter. This `mean-time0` is basically the mean of the temporal PDF of the scintillation events, in which the times are associated with the photomultiplier reference system.

This finding also clarified why α/β discrimination was more efficient in CTF. In practice, since the CTF detector was only a few metres in radius, there was basically no bias due to off-center event reconstruction. This is also confirmed using events located in a region very close to the center of Borexino, where the Gatti parameter does not show a substantial bias and exhibits a very high efficiency.

A further improvement is achieved in MLPs in which the ten `t2t`'s input variables are replaced with the ten PDF quantiles. The decision to divide the scintillation PDF into 10 intervals was found to be locally stable compared to other possible binning choices. Quantiles give the same statistical weight to the input variables, defined as one tenth of the PDF area. This definition avoids the numerical quantization problem of `t2t`'s, coming from the integer definitions of t_0's (see Fig. 7) and, more importantly, removes the correlation present by definition on the `t2t` inputs, due to the partial overlap of the integrals for different t_0.

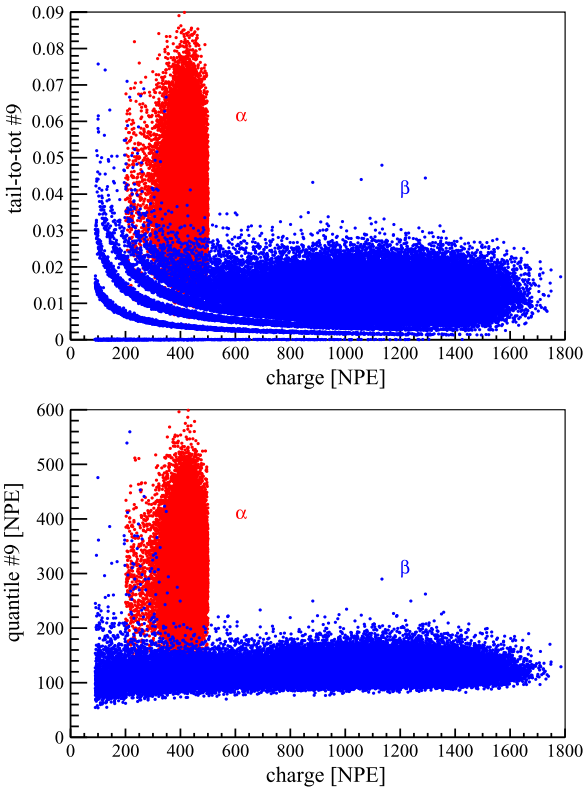

FIG. 7. Limitations of the `t2t` input variable in comparison with quantiles. Top: last `t2t` ($t > 310$ [ns]) as a function of the energy in NPE for α's (red) and β's distribution. Bottom: same distributions for the last quantile (10% tail of time PDF). The repeated pattern in the blue distribution of the top plot is related to a quantization problem arising from the choice of fixed t_0's and correlations of `t2t`'s. This problem is solved by choosing quantiles (bottom).

D. MLP test and training from ^{222}Rn events

The Gatti parameter in Borexino was initially tuned on 7000 ^{214}Bi-^{214}Po events collected during scintillator operations, before the official start of data acquisition in mid-2007. Subsequently, during six cycles of the WE campaign, between 2010 and 2011, a larger sample of ^{214}Bi-^{214}Po events was collected. This sample, on which the Gatti parameter was upgraded and the MLP studies are based, contains 85000 events in total, of which 27000 lie in the fiducial volume region $r \lesssim 3$ m. It should be noted that radon events were observed mainly at the top of the detector, in the region above the equator ($z > 0$). This evidence is also supported by the fluid dynamic simulations performed for the CNO analysis [15] and is a consequence of the effective separation between the two hemispheres due to the fluid movement in relation to the spherical geometry. After the MLP training, a slight top-bottom

asymmetry was observed, but was found to be of no practical relevance inside the analysis fiducial volume.

The final training sample (`Sample-WE`) for each MLP version contains 25000 events with $r < 3$ m from the WE period. The comparison of performances for all MLP versions and the Gatti parameter was done on a reduced sample made of about 12500 events for training, and 12500 events for test (`Sample-WEX`), both with larger radii to study also the radial dependence. Another test sample (`Sample-Ph23`), to further check the evolution of the efficiency in time, space and energy, was chosen not from ^{214}Bi-^{214}Po (basically absent after WE), but from two energy intervals of the Borexino spectrum in the first 1000 days of phase II, in which the contribution of the ^{210}Po activity is still sizeable. The α sample is selected in a very narrow region of the ^{210}Po peak (209-210 NPE), with a very small contamination of the underlying β-like component from solar neutrino and β decays; whereas the β sample is selected in the ^{7}Be shoulder region (320-400 NPE), where the leakage of the α events from the ^{210}Po right tail is also negligible.

As expected, the MLP demonstrates a decline in performance following the WE period because of PMT loss and the consequent decrease in event reconstruction resolution, leading to a dependency on both time and space (radial). Furthermore, the ^{214}Po emits a mono-energetic alpha line about 50% higher than the ^{210}Po peak energy, that falls in the region of interest for the solar neutrino analysis. As a consequence of the energy dependence of the scintillation temporal PDFs, the MLP efficiency evaluated at the ^{214}Po line is not directly applicable for the ^{210}Po analysis. The correct assessment of the space, time, and energy dependence of the MLP was studied using calibration data and Monte Carlo simulations. In the following paragraph we will report the main MLP features investigated in the Borexino analysis.

E. Different α/β MLP classifiers

The key MLP classifiers (referred to as *versions* in the Borexino terminology) that exhibited comparable performance and were utilized in the primary Borexino investigations are outlined below:

(1) `MLPv8`: This is the first version that showed a significant improvement with respect to Gatti. The input variables are the ten `t2t`'s described above, in addition to the `RMS`, the `kurtosis`, and the `mean-time0` of the non-reconstructed cluster.
(2) `MLPv10`: This version is similar to `MLPv8`, but `t2t`'s are replaced with 10 quantiles. In some cases, this version shows a slightly better performance compared to `MLPv8` due to the reasons described above, especially for low energy events.
(3) `MLPv12`: This version was meant to solve problems arising from the energy difference between the training ^{214}Bi sample, and the low energy region where the α/β discrimination is actually applied. The sample energy of ^{210}Bi is artificially reduced by reducing the number of photoelectrons (randomly removed), in a ratio of 1:2 for the ^{214}Po and of 1:4 for ^{214}Bi. Although this method assigns the correct statistical weight to the training samples (and similar to the low energy region), it cannot include real energy dependence of the scintillation PDFs.
(4) `MLPv14`: Finally, this version attempts to solve the problem of the low energy extrapolation using ^{218}Po events (with lower alpha energy and then closer to the ^{210}Po peak). Since ^{218}Po precedes the ^{214}Bi-^{214}Po fast coincidence by about 30 minutes, a space-time cut (1 meter radius and 1.5 hours before) was able to select a pure sample of about 1800 candidates. Due to the very low efficiency of the ^{218}Po tagging, this sample has limited statistics despite being a suitable representative set of low energy alpha events.

Several studies have been conducted to compare and contrast various versions of MLPs, focusing particularly on efficiency, spatial uniformity, and temporal stability.

The TMVA package returns normalized scores in the 0–1 interval, sharply peaked at 0 for α's and at 1 for β's in the Borexino choice. Figure 8 shows the distribution of the 0–1 score for the versions of interest for α's (red) and β's (blue) from test `Sample-WEX`. All of them are comparable, even if `MLPv8` shows in general better symmetry and sharper distributions. `MLPv12`, for the reasons discussed above, shows a more smeared distribution, even with a good separation. These differences can be understood by comparing the same discriminator with the plot of *receiver operating characteristic* (ROC) as reported in Fig. 9. Specifically, considering α as the signal and β as the background (note: this is opposite to the usual convention in TMVA for this specific analysis), the ROC curve reports the *true positive rate* (TPR) as a function of the *false positive rate* (FPR), changing the selection threshold m_0 in the interval $0 < m_0 < 1$, that is,

$$\mathrm{TPR} = \int_0^{m_0} \mathcal{M}_\alpha(t)dt, \tag{10}$$

$$\mathrm{FPR} = \int_0^{m_0} \mathcal{M}_\beta(t)dt, \tag{11}$$

where $\mathcal{M}_\alpha$ and $\mathcal{M}_\beta$ are the corresponding PDFs of the MLP parameters, always from the test `Sample-WEX`. From Fig. 9, one can compare the overall performance of different MLP versions (bluish and greenish curves) and the Gatti parameter (red), as reported in the figure's legend. The better discriminator approaches a right angle in the top-left corner. The figure clearly demonstrates that all MLP variants excel significantly, outperforming the Gatti method when FPR $< 0.2\%$. Specifically, for a TPR of 99.75%, it is

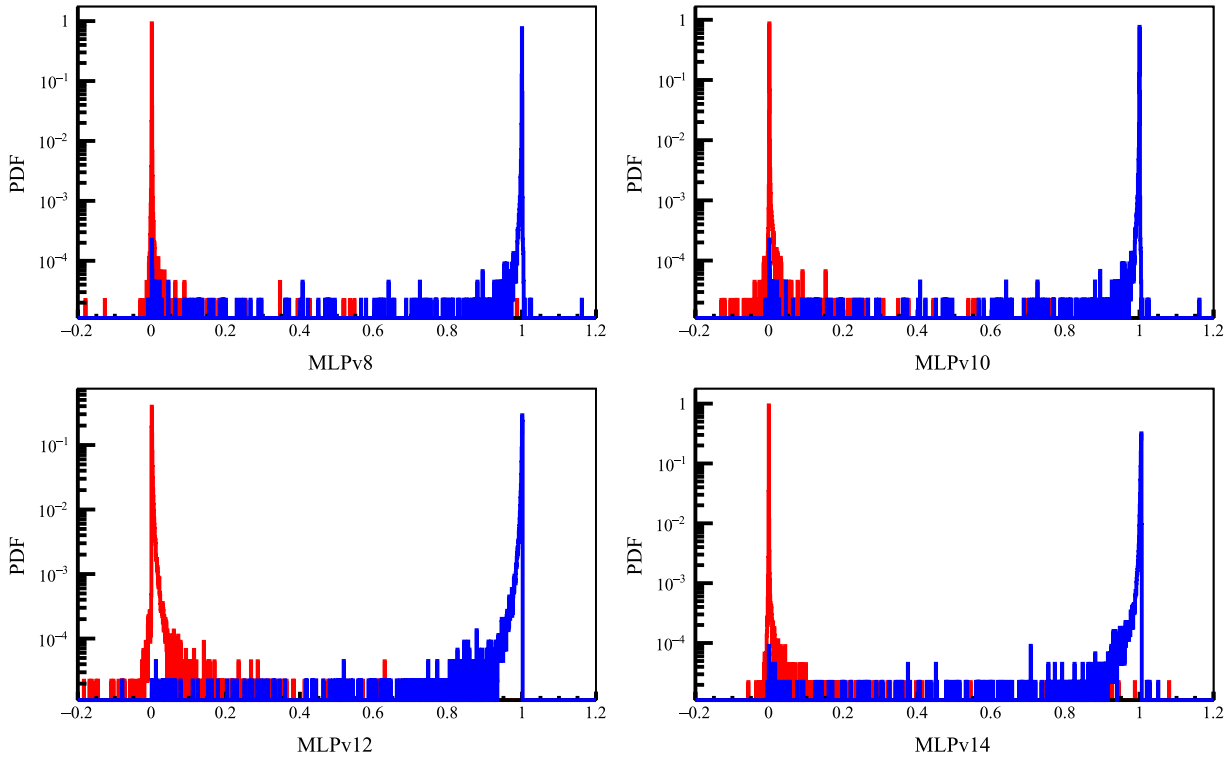

FIG. 8. Comparison between MLP versions with test `Sample-WEX`. From the top left: `MLPv8`, `MLPv10`, `MLPv12` and `MLPv14`. Red and blue PDFs represent α's and β's events, respectively.

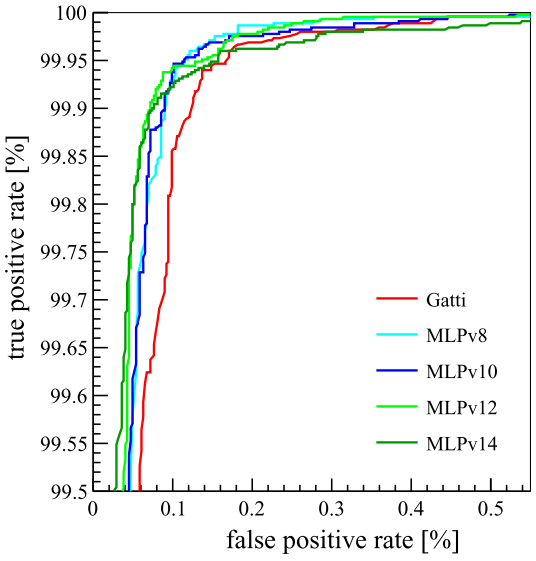

FIG. 9. Zoom of the ROC curves around the region of interest for Gatti and four MLP versions according to the color legend inset. Analysis is performed on the `Sample-WEX` test sample.

possible to achieve a twofold reduction in FPR contamination compared to the Gatti approach.

In the present study, the input parameter selection is strategically emphasized, given its pivotal role in optimizing the network's performance for the specific application. Using a two-layer network, significant separation with an exceptionally low false positive rate (only a few out of 27000 events) is successfully achieved, highlighting the efficacy of this approach despite the inherent complexities of the task, such as the initial use of a training sample from ^{214}Po for the intended ^{210}Po application. This focus was further validated by identifying a crucial parameter, the mean time for hits in the PMT reference frame, that significantly improved the event selection. These results underscore the practicality and sufficiency of a streamlined network architecture in achieving high discrimination accuracy, especially when dealing with tasks that require discrimination between different probability density functions under conditions of limited sample statistics. The success of this methodology not only meets the requirements for precise estimation in the measurement of CNO neutrinos, but also illustrates the importance of targeting parameter optimization in maximizing the potential of ANN-based applications. The selection of a two-layer method is based on a deep understanding of the limitations inherent in Borexino's experimental data.

F. MLP radial dependency

The radial dependence of the MLP score, strictly related to the position dependence of the reconstructed cluster, plays a crucial role in the ^{210}Po spatial analysis, as described at the end of Sec. III. Therefore, it is important

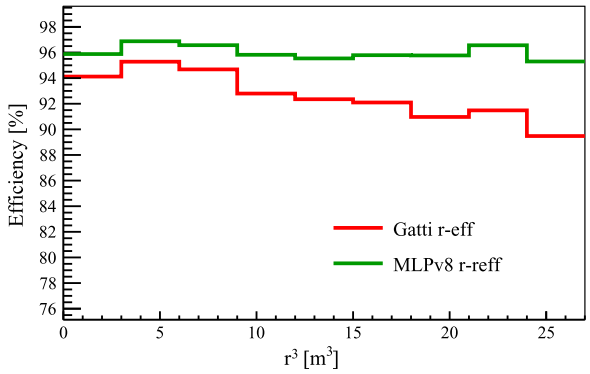

FIG. 10. Comparison between the radial efficiencies of the PSD cut for `MLPv8` < 0.3 (green) and Gatti > 0 (red). In order to account for the cubic increase in the statistics for the radial distribution, the efficiency is reported as a function of r^3, instead of r.

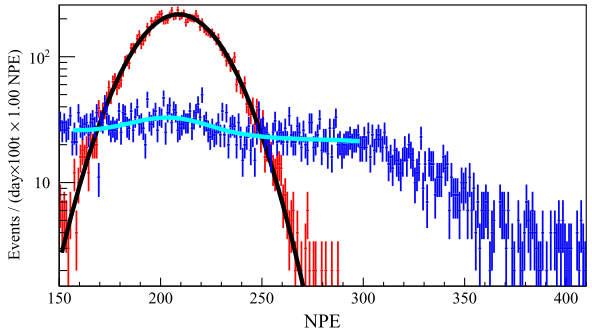

FIG. 11. Example of "MLP-subtracted" (red) and "MLP-complementary" (blue) histogram fitted with the MLP-complementary method. The red and blue distributions represent α and β complementary sets, whereas the black and cyan curves represent the fit models accounting for the exponential dependence of the MLP cut as in Eq. (12). This analysis is performed over one-year period of phase II, using `MLPv8`.

to study, through Monte Carlo simulations and test samples, any possible feature of the MLP related to the position and its possible bias in the ^{210}Po activity determination. Figure 10 shows the radial efficiency determined from the test `Sample-Ph23` for `MLPv8` (green) and Gatti (red): the first shows better behavior in terms of spatial uniformity. If one considers the CNO fiducial volume, located at about 2.75 m (about 21 m^3 on the x–axis), the efficiency is pretty uniform. As discussed in [15], the non-uniformity of the `MLPv8` efficiency is indeed negligible, as discussed in the following. Various studies have been conducted using diverse methods and varying MLP selection threshold values.

G. Stability and energy dependence of MLPs

The time dependence of the MLP, for a given selection cut, in the ^{210}Po energy region has been carefully studied for the time stability of the ^{210}Bi and ^{210}Po activity in the context of the CNO neutrino analysis; see Sec. V. In order to obtain the selection efficiency of ^{210}Po and the corresponding leakage of β events by the cut itself, events in the fiducial volume analysis are fitted with the so-called *MLP-complementary* method. In the latter, the dataset from 2011, divided yearly, are split into two histograms depending on whether the events passed the MLP cut or not, named "MLP-subtracted" and "MLP-complementary", respectively, as reported in Fig. 11.

For typical analysis, a PSD threshold of $m_0 < 0.05$ and in the 150–300 PE energy interval are used. Under these conditions, the fitted spectra as a function of the energy E [NPE] are defined as

$$S_\beta(E) = S_{\rm bx}(E)(1 - Ae^{-E/E_0}), \tag{12}$$

$$S_\alpha(E) = S_{\rm bx}(E)(Ae^{-E/E_0}), \tag{13}$$

where $S_{\rm bx}(E)$ is the typical Borexino spectrum with all fitted species (see e.g. [13]), $S_{\alpha,\beta}$ are the resulting α and β selected spectra and, finally, A and E_0 are two free parameters. Notice that the ansatz that the energy dependence of the MLP cut is exponential is suggested by Monte Carlo simulations and calibration data, and also by general considerations about the statistical nature of the neural network output. Figures 12 and 13 show the exponential energy dependence of the MLP cut from the Monte Carlo simulation and from calibration data, respectively. In both figures, the energy estimator is `npmt`, i.e. the number of hit PMTs during a scintillation event without double counting piled-up events; see [1] for further details. Both in the calibration data and in MC events, the percentages left after the MLP cut show exponential behavior, supporting the choice of the energy dependence of the efficiency assumed in the MLP-complementary fit.

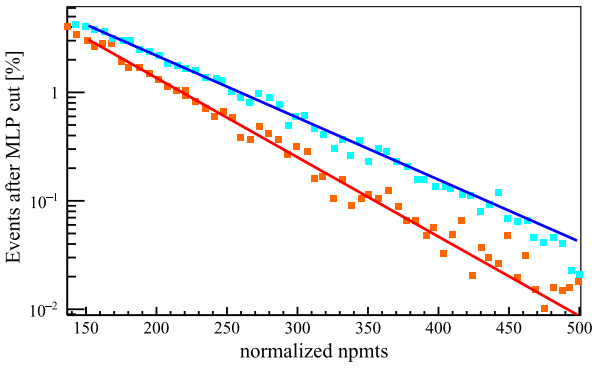

FIG. 12. Example of energy dependence (`npmt` energy estimator) of the MLP efficiency, performed over Monte Carlo data. The β-like event percentage left after the MLP < 0.015 cut are reported for the versions `MLPv12` (orange) and `MLPv10` (cyan). Both curves are fitted to exponential functions, namely the red and the blue lines in the logarithmic plot.

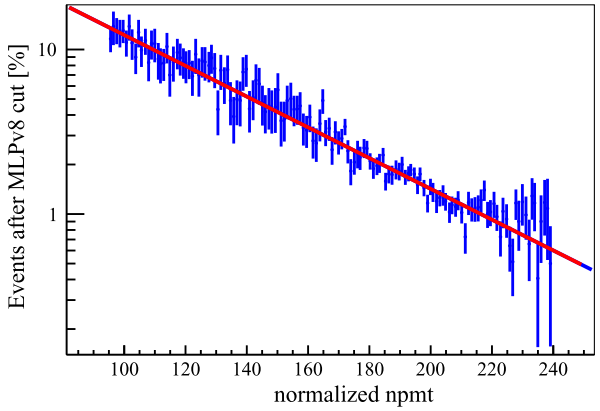

FIG. 13. Energy dependence (`npmt` estimator [1]) of the fraction of β events left after the `MLPv10` cut on real β-like calibration data (blue) fitted to an exponential curve (red).

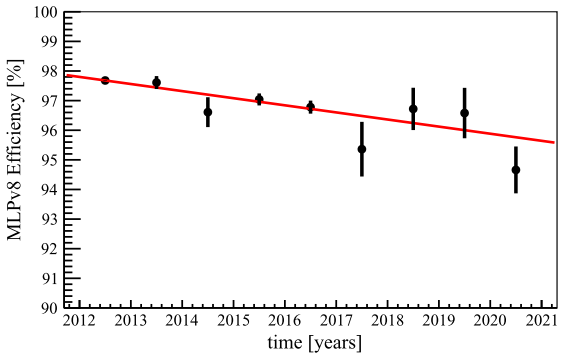

FIG. 14. MLP time dependence efficiency from MLP-complementary fit for the combined phase II and phase III period for `MLPv8`, performed over one-year time intervals (blue histogram). A fitted linear trend (red) shows, at leading order, the degradation of the MLP efficiency due to the PMTs loss (p-value $= 0.48$). The uncertainty increases over time because the reduced ^{210}Po statistics due to its decay.

Finally, Fig. 14 shows the time evolution of MLP efficiency for the ^{210}Po energy range year by year, resulting from the model in Eq. (12). The slightly decreasing linear trend is compatible with the expectation of the event reconstruction degradation, mainly related to the linear PMT loss. This dependence was used to correct the measurement of the ^{210}Po activity and to determine the corresponding systematic uncertainty for the final result on the CNO neutrino interaction rate.

V. EVENT-BY-EVENT MLP TAGGING IN THE MAIN BOREXINO ANALYSES

A. Polonium-210 studies for the CNO quest

The possibility of tagging α events with high efficiency in space and time was of crucial importance for the first measurement of neutrinos from the CNO cycle [15] with Borexino and its subsequent update [16].

Given the degeneracy and then the correlation between ^{210}Bi, pep and CNO spectra, the sensitivity to CNO neutrinos from the spectral analysis is pretty poor, unless the ^{210}Bi and pep-ν rates are independently constrained [30]. The pep-ν rate can be constrained to 1.4% precision [30], using: solar luminosity along with robust assumptions on the pp to pep neutrino rate ratio, global analysis of existing solar neutrino data [31,32], and the most recent oscillation parameters [33]. The pep constraint is essentially independent of any reasonable assumption on the CNO rate, as the solar luminosity depends only weakly on the contribution of the CNO cycle itself.

In practice, the only crucial element at play is the ^{210}Bi rate, a β emitter with a short half-life (5 days) coming from the ^{210}Pb (present in the scintillator at the beginning of phase II) through the decay chain:

$$^{210}\mathrm{Pb} \xrightarrow[22.3\ \mathrm{years}]{\beta^-} {}^{210}\mathrm{Bi} \xrightarrow[5\ \mathrm{days}]{\beta^-} {}^{210}\mathrm{Po} \xrightarrow[138.4\ \mathrm{days}]{\alpha} {}^{206}\mathrm{Pb}. \quad (14)$$

Assuming secular equilibrium, the ^{210}Bi rate can be determined from the ^{210}Po activity [30,34,35]. Since ^{210}Po activity can be measured precisely by highly efficient MLP α/β tagging, this strategy provided the key solution to tackle species correlation in the spectral fit and led to the first observation of the CNO neutrino interaction rate in Borexino and its subsequent upgrades [15,16]. At the beginning of Borexino phase II (early 2013), it was clear that the presence of convection motions, caused by the seasonal change of the temperature in the Gran Sasso experimental hall, made it impossible to exploit the ^{210}Bi-^{210}Po link, as suggested by the sequence in Eq. (14). To solve this problem, a long and challenging thermal stabilization program was undertaken by the collaboration to prevent the scintillation convective motion, and the consequent contaminant mixing in the scintillator. This program, started in mid-2014, consisted of different phases: (i) installation of high precision temperature probes inside and outside the detector, (ii) thermal insulation of the detector with different layers of rock wool, (iii) active temperature control systems of the detector and (iv) of the experimental room. This long-standing effort worked successfully and allowed Borexino to set an upper limit on the ^{210}Bi rate, a crucial ingredient for the final extraction of the CNO neutrino interaction rate from the spectral analysis.

It is worth mentioning that the MLP tagging with its high efficiency uniformity and stability helped in all stages of this enterprise: from the understanding of the ^{210}Po migration in the scintillator, through the study of the effects of the different phases of the thermal insulation program, to the determination of the ^{210}Bi upper limit rate (see for details the Appendix of [15]).

For the CNO analysis, the space and time dependence of the α/β tag in the ^{210}Po region was carefully studied using the MLP complementary analysis and Monte Carlo simulations. In particular, the latter was crucial for the optimization of the cut and for the evaluation of the efficiency dependence on the radial position and time. In particular the best cut was defined by maximizing the standard signal-to-background (S/B) figure of merit (FoM):

$$\mathrm{FoM} = \frac{S}{\sqrt{S+B}}. \tag{15}$$

In this case S can be assumed as true positive events (real α's), and B as false positive events (β-like events leaking out from the distribution tail). For `MLPv8` the best α cut, corresponding to the phase III dataset, was found at $m_0 < 0.3$.

B. MLP in other analyses

In addition to CNO analysis, MLP α/β tagging was used in many other analyses published by the Borexino Collaboration. In particular, it played an important role in the high-importance detection of the seasonal modulation of ^{7}Be neutrinos due to the eccentricity of the Earth orbit [36], for the reduction of the ^{210}Po component in the region of interest of the ^{7}Be spectrum. This analysis was updated including the entire phase II and phase III dataset, leading to the first independent measurement of the Earth orbit eccentricity with only solar neutrinos [37].

In the geo-neutrino analysis, the MLP was used in event selection with high performance even at large radii (∼4 m) close to the Nylon vessel [23] and for the ^{210}Po background estimation for the neutron background induced by alpha decays. In addition, the MLP selection was used for the space and time selection of the ^{210}Po data events for the accurate Monte Carlo tuning used to simulate the ^{210}Po spectrum in phase II [38] and phase III. In particular, this study played an important role in the comprehensive analysis of the pp chain [12,13]. This study has provided a measurement of the most important solar neutrino fluxes, which is in favor of the MSW-LMA neutrino oscillation scenario at 98% CL. (see [39] and Refs. therein).

VI. CONCLUSIONS

In this paper, we offer a detailed review of the α/β pulse shape discrimination adopted in Borexino. We present various implementations used during more than a decade of data taking, starting with the Gatti optimal filter and the corresponding statistical subtraction of the α component from the energy spectrum, ending with the more sophisticated PSDs based on ANNs, specifically exploiting MLP. The latter, with its high efficiency, spatial uniformity, and time stability, allowed us to event-by-event select the ^{210}Po events, a crucially important background reduction which made possible the observation of CNO solar neutrinos in Borexino.

Compared to the Gatti parameter approach, ANNs single out parameters relevant to PSD in a highly non-factorisable way. In the case of Borexino, the Gatti parameter was limited by information loss in the photon arrival times after position reconstruction correction. Integration of variables into the MLP before event reconstruction improved the performance of the α/β selection. The MLP implementation required careful calibration to select the best input parameters, adjust the algorithm, and evaluate its performance. Its spatial and time efficiency were monitored and used for the evaluation of the global systematic uncertainty of some of the most important Borexino results for which the method was used.

The α/β pulse shape discrimination allowed by intrinsic properties of the scintillator, was proven to be fully exploitable in an ultrapure, large-volume detector such as Borexino. In particular, it played an essential role in the neutrino spectroscopy of the entire pp chain and the first observation of neutrinos from the CNO cycle.

ACKNOWLEDGMENTS

We acknowledge the generous hospitality and support of the Laboratori Nazionali del Gran Sasso (Italy). The Borexino program is made possible by funding from Istituto Nazionale di Fisica Nucleare (INFN) (Italy), National Science Foundation (NSF) (USA), Deutsche Forschungsgemeinschaft (DFG), Cluster of Excellence PRISMA+ (Project ID 39083149), and recruitment initiative of Helmholtz-Gemeinschaft (HGF) (Germany), Russian Science Foundation (RSF) (Grant No. 24-12-00046) and Ministry of Science and Higher Education of the Russian Federation (Project FSWU-2023-0073) (Russia), and Narodowe Centrum Nauki (NCN) (Grant No. UMO-2013/10/E/ST2/00180) (Poland). We gratefully acknowledge the computing services of Bologna INFN-CNAF data center and U-Lite Computing Center and Network Service at LNGS (Italy). This research was supported in part by PLGrid Infrastructure (Poland).

[1] G. Bellini *et al.* (Borexino Collaboration), Phys. Rev. D **89**, 112007 (2014).
[2] G. Bellini *et al.* (Borexino Collaboration), Phys. Rev. Lett. **108**, 051302 (2012).
[3] M. Agostini *et al.* (Borexino Collaboration), Nature (London) **562**, 505 (2018).
[4] G. Ranucci, R. Dossi, P. Inzani, G. Korga, P. Lombardi, E. Meroni, and M.E. Monzani, IEEE Trans. Nucl. Sci. **51**, 1784 (2004).
[5] H. O. Back *et al.* (Borexino Collaboration), Nucl. Instrum. Methods Phys. Res., Sect. A **584**, 98 (2008).
[6] E. Gatti and F. De Martini, *Proceedings of the IAEA, Wien* (1962), Vol. 2, pp. 265–276, https://inis.iaea.org/collection/NCLCollectionStore/_Public/43/116/43116625.pdf.
[7] C. Arpesella *et al.* (Borexino Collaboration), Phys. Lett. B **658**, 101 (2008).
[8] C. Arpesella *et al.* (Borexino Collaboration), Phys. Rev. Lett. **101**, 091302 (2008).
[9] G. Bellini *et al.* (Boreino Collaboration), Phys. Rev. Lett. **107**, 141302 (2011).
[10] Website: https://www.lngs.infn.it.
[11] G. Alimonti *et al.* (Borexino Collaboration), Astropart. Phys. **16**, 205 (2002).
[12] G. Bellini *et al.* (Borexino Collaboration), Nature (London) **512**, 383 (2014).
[13] M. Agostini *et al.* (Borexino Collaboration), Phys. Rev. D **100**, 082004 (2019).
[14] M. Agostini *et al.* (Borexino Collaboration), Phys. Rev. D **101**, 062001 (2020).
[15] M. Agostini *et al.* (Borexino Collaboration), Nature (London) **587**, 577 (2020).
[16] S. Appel *et al.* (Borexino Collaboration), Phys. Rev. Lett. **129**, 252701 (2022).
[17] D. Basilico *et al.* (Borexino Collaboration), Phys. Rev. D **108**, 102005 (2023).
[18] A. Vishneva *et al.* (Borexino Collaboration), J. Phys. Conf. Ser. **888**, 012193 (2017).
[19] S. K. Agarwalla *et al.* (Borexino Collaboration), J. High Energy Phys. 02 (2020) 038.
[20] M. Agostini *et al.* (Borexino Collaboration), Astropart. Phys. **125**, 102509 (2021).
[21] M. Agostini *et al.* (Borexino Collaboration), Phys. Rev. D **96**, 091103 (2017).
[22] G. Bellini *et al.* (Borexino Collaboration), Phys. Rev. D **88**, 072010 (2013).
[23] M. Agostini *et al.* (Borexino Collaboration), Phys. Rev. D **101**, 012009 (2020).
[24] C. Galbiati and K. McCarty, Nucl. Instrum. Methods Phys. Res., Sect. A **568**, 700 (2006).
[25] M. J. Carvalho and G. Klein, J. Lumin. **18**, 487 (1979).
[26] G. Ranucci, A. Goretti, and P. Lombardi, Nucl. Instrum. Methods Phys. Res., Sect. A **412**, 374 (1998).
[27] G. Alimonti *et al.* (Borexino Collaboration), Phys. Lett. B **422**, 349 (1998).
[28] J. J. Hopfield, Proc. Natl. Acad. Sci. U.S.A. **79**, 2554 (1982).
[29] A. Hocker, P. Speckmayer, J. Stelzer, J. Therhaag, E. von Toerne, H. Voss, M. Backes, T. Carli, O. Cohen, A. Christov *et al.*, arXiv:physics/0703039.
[30] M. Agostini *et al.* (Borexino Collaboration), Eur. Phys. J. C **80**, 1091 (2020).
[31] F. Vissani, in *Solar Neutrinos*, edited by M. Meyer (World Scientific, Singapore, 2019), pp. 121–141.
[32] J. Bergström, M. C. Gonzalez-Garcia, M. Maltoni, C. Peña-Garay, A. M. Serenelli, and N. Song, J. High Energy Phys. 03 (2016) 132.
[33] F. Capozzi, E. Lisi, A. Marrone, and A. Palazzo, J. Phys. Conf. Ser. **1312**, 012005 (2019).
[34] E. Rutherford, *Radio-Activity* (Harward University Press, 1905), p. 331; H. Bateman, Proc. Cambridge Philos. Soc. **15**, 423 (1910).
[35] F. L. Villante, A. Ianni, F. Lombardi, G. Pagliaroli, and F. Vissani, Phys. Lett. B **701**, 336 (2011).
[36] M. Agostini *et al.* (Borexino Collaboration), Astropart. Phys. **92**, 21 (2017).
[37] S. Appel *et al.* (Borexino Collaboration), Astropart. Phys. **145**, 102778 (2023).
[38] M. Agostini *et al.* (Borexino Collaboration), Astropart. Phys. **97**, 136 (2018).
[39] S. T. Petcov and M. Piai, Phys. Lett. B **533**, 94 (2002).
[40] H. O. Back *et al.* (Borexino Collaboration), J. Instrum. **7**, P10018 (2012).