A Cross-correlation Study of High-energy Neutrinos and Tracers of Large-scale Structure

Ke Fang1,3, Arka Banerjee1,2, Eric Charles1,2, and Yuuki Omori1

1 Kavli Institute for Particle Astrophysics and Cosmology (KIPAC), Stanford University, Stanford, CA 94305, USA
2 SLAC National Accelerator Laboratory, 2575 Sand Hill Road, Menlo Park, CA 94025, USA

Received 2020 February 22; revised 2020 March 21; accepted 2020 March 30; published 2020 May 12

Abstract

The origin of the bulk of the astrophysical neutrinos detected by the IceCube Observatory remains a mystery. Previous source-finding analyses compared the directions of IceCube events and individual sources in astrophysical catalogs. The source association method is technically challenging when the number of source candidates is much larger than the number of the observed astrophysical neutrinos. We show that in this large source number regime, a two-point cross-correlation analysis of neutrino data and source catalog can instead be used to constrain potential source populations for the high-energy astrophysical neutrinos, and provide spatial evidence for the existence of astrophysical neutrinos. We present an analysis of the cross-correlation of the IceCube 2010–2012 point-source data and a WISE–2MASS galaxy sample. While we find no significant detection of cross-correlation with the publicly available neutrino data set, we show that, when applied to the full IceCube data, which has a longer observation time and higher astrophysical neutrino purity, our method has sufficient statistical power to detect a cross-correlation signal if the neutrino sources trace the large-scale structure of the universe.

Unified Astronomy Thesaurus concepts: Neutrino astronomy (1100); Large-scale structure of the universe (902)

1. Introduction

The existence of an astrophysical neutrino population above $\sim$3–10 TeV has been established by the IceCube Observatory (Aartsen et al. 2013, 2016; Schneider 2019; Stettner 2019). The discovery is based on an excess of the observed flux over the atmospheric background in several detection channels. No signatures of clustering of astrophysical neutrinos have been measured. The origin of these cosmic neutrinos remains unknown. Point-source searches have been carried out, in the form of blind searches that scan the sky with subdegree grids (Aartsen et al. 2018, 2019b, 2020a), and source association searches that stack the likelihoods of sources in a given catalog (e.g., Aartsen et al. 2017a, 2019a; Mertsch et al. 2017). Note that the latter are sometimes referred as “correlation” analyses in the literature. To avoid confusion we specifically call them “likelihood stacking analyses” below. None of the searches has led to a robust identification of the sources of the bulk of the observed neutrinos, though hints of sources have been suggested (Aartsen et al. 2020b).

Formally, the likelihood stacking analysis requires computation of the probability of each neutrino coming from every source in the galaxy catalog. The complexity of such an analysis scales with the number of sources in a catalog, and can be computationally challenging due to factors such as spatial correlations of closely spaced sources. However, high-energy neutrinos may plausibly come from sources from a large population with a number density $\geq 10^{-6}$ Mpc$^{-3}$ (more than a million sources within redshift $z \sim 1$), such as star formation activities and their host galaxies (e.g., Fang et al. 2014; Tamborra et al. 2014; Bechtol et al. 2017) and galaxy clusters (e.g., Fang & Murase 2018). In this paper, we present a novel approach to test the connection of IceCube events with potential sources that are large in population.

The two-point correlation function is a common statistical tool used to describe the distribution of galaxies (Peebles 1980). It is a measure of the excess probability of finding a pair of data points at some separation $R$ compared to a random distribution of points. The cross-correlation generalizes this definition to the case when the two data points are drawn from different data sets. The two-point cross-correlation function has been used in analyzing $\gamma$-ray data (e.g., Ando et al. 2015; Feyereisen et al. 2017; Ammazzalorso et al. 2019). In this work, we show that the cross-correlation function, or its equivalent in harmonic space, the cross power spectrum, can be applied to the study of high-energy neutrinos. We have developed the framework and an analysis pipeline to compute the cross-correlation between IceCube events and galaxy samples and applied this methodology to the IceCube 2010–2012 point-source data set and a galaxy catalog based on the Wide-Field Infrared Survey Explorer (WISE; Wright et al. 2010) and the 2-Micron All-Sky Survey (2MASS; Skrutskie et al. 2006) infrared databases. Unlike the stacking analyses that were applied to test, e.g., finding whether neutrinos come from sources in the Fermi-2LAC Blazar catalog (Aartsen et al. 2017a), the cross-correlation analysis outlined here does not scale in computational cost with the number of sources, and does not require the use of a catalog of the exact neutrino source class, as long as neutrino sources and the galaxy sample used in the analysis both trace the same underlying large-scale density field. Detection of cross-correlations with different galaxy samples covering different galaxy types and redshift distributions could potentially narrow down the source classes and redshift range of origin for the high-energy neutrinos.

The paper is organized as follows. The method is laid out in Section 2 and Appendix A. The analysis is described in Section 3 and Appendices A.1, A.2. C. Findings are summarized in Section 4 and discussion and conclusions are presented in Section 5.
2. Cross-correlation of Neutrinos and Galaxies

Let \( n_1(x) \) and \( n_2(x) \) represent the number density of galaxies and the number density of neutrino events detected by IceCube at some position \( x \) on the plane of the sky respectively. For both fields, we can define the overdensity field as

\[
\delta(x) = \frac{n(x) - \bar{n}}{\bar{n}},
\]

where \( \bar{n} \) is the average value of \( n(x) \) over all directions used in the analysis. Below, we refer to \( \delta \) as the fluctuation or overdensity field interchangeably.

The two-point cross-correlation function of the galaxy and neutrino fluctuation fields, and \( C_{\ell}^{gg} \), is defined as

\[
C_{\ell}^{gg} = \frac{4\pi}{2\ell + 1} \int d\cos \theta \langle \delta_1(\mathbf{x}) \delta_2(\mathbf{x'}) \rangle P_{\ell}^g(\cos \theta),
\]

where \( \theta \) is the angle between the two directions \( x \) and \( x' \), \( P_{\ell} \) is the Legendre polynomial, and the average is performed in the \( x \) space. The term \( C_{\ell}^{gg} \) denotes the contribution of the harmonic \( \ell \) to the correlation function,

\[
C_{\ell}^{gg} = \frac{1}{f_{\text{sky}}(2\ell + 1)} \sum_{m=-\ell}^{\ell} a_{\ell m}^{gg} a_{\ell m}^{\ast},
\]

where \( a_{\ell m} \) are the coefficients to decompose the fluctuation field into spherical harmonics, \( \delta(\theta, \phi) = \sum_{\ell m=0}^{\infty} a_{\ell m} Y_{\ell m}(\theta, \phi) \).

Assuming that neutrino sources, denoted as \( s \), such as blue star-forming galaxies, trace the same underlying matter density modes as the galaxies used in the analysis, denoted as \( g \). On large scales (\( \ell \ll 300 \)) we can write (Bardeen et al. 1986)

\[
C_{\ell}^{gg} = b_g^2 C_{\ell}^{mm},
\]

\[
C_{\ell}^{gg} = b_s^2 C_{\ell}^{mm},
\]

where \( b_g, b_s \) are the bias parameters of the two populations, and \( C_{\ell}^{mm} \) represents the power spectrum of the underlying density field. If the astrophysical neutrinos detected by IceCube represent a Poisson sampling from the source population, then the cross-correlation spectrum of these events with the galaxy catalog can be written as

\[
C_{\ell}^{qf} = f_g C_{\ell}^{gg}.
\]

Comparing Equation (5) to (4), we see that \( f_g = b_s/b_g \). The value of \( f_g \) depends on value of the bias of the galaxy sample used (\( b_g \)), as well as the source population (\( b_s \)). However, for most galaxy samples, and for most of the proposed source populations, the bias values are \( \sim 0.1 \), and therefore \( f_g \) is also expected to be \( \mathcal{O}(1) \). For example, if we use BOSS–CMASS galaxies around \( z \sim 0.5 \) as the galaxy sample, \( b_g \sim 2 \) (Nuza et al. 2013), and if the source population is blue star-forming galaxies in a similar redshift range, \( b_s \sim 1 \) (e.g., López-Sanjuan et al. 2017), implying \( f_g \sim 0.5 \).

---

4 The bias parameter relates the clustering of the peaks of a Gaussian random field to the clustering of the underlying field (Kaiser 1984). More massive halos generally form on rarer peaks of the initial field, and have a higher bias parameter. \( b_g \) is the bias parameter of the galaxies that source the neutrinos.

It should be noted that the IceCube data are dominated by atmospheric neutrinos up to \( \sim 100 \) TeV in the northern sky, and by atmospheric muons up to the highest energies in the southern sky. For a mixed population of astrophysical and atmospheric neutrinos from cosmic-ray interactions, the cross-power spectrum satisfies

\[
C_{\ell}^{gg} = f_{\text{astro},i} C_{\ell}^{gg} + (1 - f_{\text{astro},i}) C_{\ell}^{g \text{ atm}},
\]

with \( f_{\text{astro},i} \) being the fraction of astrophysical events in energy bin \( i \) (see Appendix A).

Assuming that the energy bins are independent (Aartsen et al. 2014), we define the likelihood function by

\[
\log L(C_{\ell}^{gg} | f_{\text{astro}}) = -\sum_{\ell i} \frac{(C_{\ell}^{gg} - (C_{\ell}^{gg}(f_{\text{astro},i})))^2}{2(\sigma_{\ell i}^g)^2}.
\]

As atmospheric neutrinos and muons do not trace the distribution of galaxies, \( (C_{\ell}^{g \text{ atm}}) = 0 \), and the expected mean cross-correlation of a combined sample of astrophysical and atmospheric neutrinos is \( (C_{\ell}^{gg}) = f_g C_{\ell}^{gg} \). The expected standard deviation, \( \sigma_{\ell i}^g \), of a combined sample is obtained by running a set of Monte Carlo simulations that contain both astrophysical and atmospheric events, and taking the standard deviation of the cross-correlation values obtained in each sample (see Appendix A.2). In general we find that \( \sigma_{\ell i}^g \) is insensitive to \( f_{\text{astro},i} \), and is inversely proportional to the square root of the sample size \( \sim N_v^{-1/2} \).

The significance of a signal against the null hypothesis, defined as zero correlation between the neutrino sample and the source catalog, can be quantified by a test statistic,

\[
\text{TS} = 2\log L(f_g, \hat{f}_{\text{astro}}) - \log L(0),
\]

where \( \hat{f}_g \) and \( \hat{f}_{\text{astro}} \) are the maximum likelihood values of \( f_g \) and \( f_{\text{astro}} \), respectively. If the neutrino and galaxy fluctuation fields are Gaussian, the TS should follow the chi-square distribution with degrees of freedom equal to the number of energy bins used for the likelihood evaluation (Wilks 1938).

3. Analysis Setup

An ideal neutrino data set for the cross-correlation study would have full-sky coverage, good angular resolution, large sample size, and high purity of astrophysical neutrino events. The veto techniques of IceCube (e.g., Aartsen et al. 2013, 2015a) reduce atmospheric muon background by selecting neutrino events that interact within the detector boundary. The through-going tracks from the northern hemisphere also provide a clean neutrino sample, as up-going muons are suppressed by the Earth (Stettner 2019). These contained-vertex and through-going track event samples are suitable for the proposed analysis.

Of the public data sets, only the point-source data set (Aartsen et al. 2017b) includes both the full direction information and the corresponding effective area tables. Therefore, we tailored the public point-source data set for a demonstration of the cross-correlation analysis.

The point-source data are composed of track-like events with angular resolutions ranging from \( \sim 1^\circ \) around 1 TeV to \( \sim 0.4^\circ \).
above 10 TeV (Aartsen et al. 2017b). We bin the data into HEALPix\(^7\) sky maps with \(N_{\text{side}} = 128\) using the HEALPy package\(^8\) (Górski et al. 2005; Zonca et al. 2019). To ensure that enough counts are available for the analysis, we group events into decade-wide logarithmic energy bins ranging from \(10^{1.5}\) to \(10^{8.5}\) GeV. To avoid the large atmospheric muon background, we only use events from the northern sky, defined as decl. angle decl. > \(-5^\circ\) (Aartsen et al. 2017b). The counts map, distribution of the zenith angle, and the autocorrelation of the point-source data are presented in Appendix A.1.

As shown in Figure B2, the spatial distribution of the effective area of the IceCube point-source data is smoother than that of three-year neutrino events on angular scales \(\ell \gtrsim 50\). Event counts thus trace the source distribution on these small angular scales. For this study we prefer counts to flux as an indicator of the source distribution, as single events in spatial bins with very low exposure can result in anomalously large fluxes.

We generate “negative control” samples of synthetic atmospheric neutrino data based on the zenith angle distribution of observed events (see Appendix A.2). To test the sensitivity of our method in finding a cross-correlation signal, we sample astrophysical events from the density field of our galaxy sample (see below) and set \(f_{\text{astro}}\) based on the diffuse muon neutrino flux in the IceCube 10-year data (Figure 3 of Stettner 2019), \(f_{\text{astro}} = f_{\text{astro}}^\text{mc}\). The sample size of the synthetic data is randomly generated from a Poisson distribution centered at the observed event count. As the IC79-2010 data have different spatial and energy distributions from the IC86-2010 and 2012 data, we generate synthetic data year by year and sum the resulting count maps.

The all-sky galaxy catalog used in this work is constructed following Kovács & Szapudi (2015) to combine photometric information of the WISE (Wright et al. 2010) and 2MASS (Skrutskie et al. 2006) infrared databases (see Appendix C for more details). The region with Galactic latitude \(||b| < 10^\circ|\) is masked to avoid Galactic foregrounds. The median redshift of the sample is \(z \approx 0.14\).

The \(a_{\ell m}\) coefficients of the neutrino and galaxy overdensities, and the \(s_{\ell k}\) accounting for neutrino and galaxy masks are used to compute \(C_{\ell}^{\nu g}\) following Equation (3). We set \(\ell_{\text{min}} = 50\) when computing the likelihood in Equation (7) to avoid effects from the nonuniform IceCube exposure and the masks of neutrino and galaxy data (see Appendix A.1). The results, however, do not significantly depend on \(\ell_{\text{min}}\) as long as \(\ell_{\text{min}} > 5\). The standard deviations of the model, \(\sigma_{\ell k}^{\nu g}\), are precomputed from 500 realizations of synthetic data. We use three energy bins \(i = 1–3\) (uniform in the logarithm from 0.3 to 300 TeV) for the likelihood calculation, considering that the data are heavily dominated by astrophysical events below \(\sim 3\) TeV, and that the three-year point-source data have no events above 3 PeV in the northern sky.

The fractions \(f_{\text{astro}}^i\) in the different energy bins \(i\) are assumed to be independent, and are coupled with \(f_\nu\). The value of \(f_\nu\) depends on the redshift of the galaxy sample and the type of neutrino source, but is generally on the order of unity (see Section 2). We thus set \(f_{\text{astro}}^1 \equiv f_\nu f_{\text{astro}}\) and allow a higher upper bound for \(f_{\text{astro},1}^\nu\) to account for the cases where \(f_\nu > 1\). We also allow negative \(f_{\text{astro}}\) to describe anticorrelation of neutrino sources and galaxy catalogs. The results are insensitive to the upper bound, and we obtain similar results with upper bounds ranging from 1 to 4.

### 4. Results

Cross-correlating the northern-sky events in the IceCube three-year point-source data and the WISE–2MASS galaxy sample leads to best-fit \(f_{\text{astro},1,2,3} = (0.011, -0.027, -0.076)\) and \(TS = 4.3\) (corresponding to 1.2σ confidence level in a two-normal distribution). We find no evidence of astrophysical neutrinos in the tailored data set that follow the spatial distribution of the WISE–2MASS galaxy sample.

We take a Bayesian approach to sample the parameter space of \(f_{\text{astro}}\) using Markov Chain Monte Carlo (MCMC) via emcee.\(^9\) We adopt a uniform prior probability for \(-4 < f_{\text{astro},i} < 4\). The left panel of Figure 1 presents the findings of an ensemble sampler with 640 walkers and 500 steps. The black contours indicate the 68%, 95%, and 99% confidence intervals for \(f_{\text{astro},1,2,3}\) accordingly.

The fraction of astrophysical events that cross correlate with galaxies can be converted into an energy flux through

\[
\phi_{\nu}(E_i) \approx \frac{N_{\nu,i} f_{\text{astro},i}}{\Delta E_i \Delta t f_{\text{sky}} 4 \pi A_{\text{eff},i}},
\]

where \(N_{\nu,i}\) is the number of neutrino events in bin \(i\), \(\Delta t\) is the active observation time of IceCube, and \(A_{\text{eff},i}\) is the mean of the weighted effective area (as defined in Equation (B1)) over the unmasked region, which accounts for the angular dependence of the effective area. To avoid negative flux, we assume \(f_\nu = 1\) and use a one-sided test \(f_{\text{astro}} > 0\) to obtain the fractions before converting them to flux. Depending on the significance of the cross-correlation, we quote \(\phi_{\nu}\) as an upper limit at the 95% C. L. (when TS < 4) or a data point with 1σ error bars (when TS > 4, corresponding to 2σ with one degree of freedom). The upper limits of \(f_{\text{astro}}\) at 95% C. L. are (0.022, 0.016, 0.16). The right panel of Figure 1 presents the converted upper limits to the energy flux of astrophysical neutrinos whose sources are distributed like the WISE–2MASS galaxy sample.

To demonstrate the efficiency of the cross-correlation method, Figure 2 presents the cumulative probability distribution of the TS of cross-correlation between the WISE–2MASS galaxy sample and the synthetic 3-year and 10-year data. The plot is made from 10\(^4\) realizations of synthetic data that contain only atmospheric neutrinos or atmospheric neutrinos plus astrophysical neutrinos sampled from the galaxy density field. The TS distributions of the background-only samples agree with the chi-square distribution, confirming the Gaussianity of our likelihood function. The filled regions denote the TS distributions of mixed-population neutrino samples from 3-year (light blue) and 10-year (dark blue) observations. The lower bounds of the filled regions correspond to a muon neutrino-like sample, with input \(f_{\text{astro}} = f_{\text{astro}}^\nu\) suggested by the diffuse muon neutrino analysis (Stettner 2019; also see Appendix A.2). The upper bounds correspond to a more optimistic scenario with \(f_{\text{astro}} = 2 f_{\text{astro}}^\nu\) motivated by the fraction of astrophysical events above 30 TeV in the high-energy starting event (HESE) sample (Schneider 2019). With 10 years of full-sky IceCube

\(^7\) http://healpix.sf.net

\(^8\) https://healpy.readthedocs.io/en/latest/

\(^9\) https://emcee.readthedocs.io/en/stable/
log-likelihood found by maximizing Equation \(\text{correlates with the overdensity limits at 95\% C.L.}\). The contours indicate the 68\%, 95\%, and 99\% C.L. regions for individual parameters found by an MCMC sampling of the parameter space. Right: upper function of a chi-square distribution with 3 degrees of freedom of observation. For comparison, the gray dashed line shows the probability indicates neutrino samples composed of only atmospheric events from 10 years values of \(f\) to the number of energy bins used for the analysis.

Further, in the scenario of complete galaxy samples and sufficient astrophysical neutrino events at IceCube, detections and nondetections of cross-correlation with each galaxy sample can be used to narrow down the sources of the IceCube neutrinos. However, we show that if the analysis is performed with the full IceCube data of contained events and northern-sky muon neutrino events, we should have a statistically significant detection of any true cross-correlation. We urge an immediate analysis follow-up by the high-energy neutrino experiments.

Apart from increasing the sample of neutrino events, improvements are also possible by optimizing the LSS tracer used in the analysis. The cross-correlation analysis is best performed using a complete and clean tracer of the LSS. The current analysis uses a sample of galaxies constructed from the WISE and 2MASS infrared surveys. The sample has low stellar contamination and high completeness, but only contains nearby galaxies with a median redshift \(z \approx 0.14\). To better search for

5. Discussion and Conclusions

High-energy neutrinos are a unique messenger of hadronic processes of the universe at extreme energies. Understanding their origin is a crucial task in neutrino astronomy. Previous source searches have focused on the association of IceCube events with individual sources in a catalog, and are thus limited to source classes with relatively small populations. In this paper, we have performed the first cross-correlation analysis between the IceCube events and a tracer of the LSS—galaxies from a WISE–2MASS catalog. A nonzero cross-correlation is expected if the source population generating high-energy neutrinos, such as star-forming galaxies and galaxy clusters, traces the same underlying matter density modes as the galaxy sample. Such a detection would be the first spatial evidence for the astrophysical origin of the high-energy neutrino events.

Using the the publicly available three-year point-source data set from IceCube we do not find a significant detection of cross-correlation with the WISE–2MASS galaxy sample. However, we show that if the analysis is performed with the full IceCube data of contained events and northern-sky muon neutrino events, we should have a statistically significant detection of any true cross-correlation. We urge an immediate analysis follow-up by the high-energy neutrino experiments.
cross-correlation with LSS, samples with higher number densities, different redshift ranges, and larger sky coverage can be used. For example, the Sloan Digital Sky Survey provides a wide-field coverage of galaxies out to $z \sim 0.8$ in the optical wavelength range (Ahumada et al. 2019). It is also possible to use tracers of LSS other than galaxies. For example, the cosmic infrared background map from Planck (Lenz et al. 2019) consists of infrared emission from dusty galaxies and traces the star formation history. Measurements of the weak lensing shear from the Dark Energy Survey (Chang et al. 2018) can also be used as a proxy for the underlying matter field on large scales. Finally, cross-correlating with samples of galaxy clusters (e.g., Bleem et al. 2015; Clerc et al. 2016) can directly answer the question whether neutrinos come from cosmic-ray interaction in the intracluster medium.

As pointed out in Section 2, $f_{\text{astro}}$ and $f_{\ell}$ are currently strongly correlated. However, we note that this degeneracy may be broken if the energy spectrum of the neutrino sample is known. Measurements of the diffuse astrophysical neutrino flux provide an estimation of $f_{\text{astro}}$ based on the modeling of the atmospheric neutrino contribution. Then the cross-correlation analysis can derive $f_{\ell}$ which would inform the relation of the neutrino sources to the test sources in use.

Our method applies when neutrino sources have a large population. A large source number makes the detection of individual sources difficult (see, e.g., Fang et al. 2016; Feyereisen et al. 2017), but enables the comparison of the distribution patterns of sources and neutrinos. The method is not optimized for a scenario where the neutrino emission is dominated by a small group of bright sources. Signals in large angular scales (below $\ell_{\text{min}}$) should be carefully treated with respect to the unevenness in the IceCube effective area. The source association method used in Aartsen et al. (2017a, 2019a) is a more natural approach in this regime.

The analysis code we developed to perform cross-correlation of IceCube events and galaxy samples is available.\textsuperscript{10} It is written in Python and uses healpy to perform calculations of spherical harmonics.

We thank Erik Blaufuss and Mike Richman for their helpful discussion about the IceCube effective area of the public point-source data. We thank Seth Digel for his extremely useful feedback on our manuscript.

Appendix A

Cross-correlations in the Presence of a Mixture of Populations

The neutrinos detected by IceCube in any given energy bin come from two distinct populations—astrophysical neutrinos and atmospheric neutrinos. The total number density of neutrinos detected at some point $n_\nu(x)$ on the sky is given by

\[ n_\nu(x) = n_{\text{astro}}(x) + n_{\text{atm}}(x). \]  

(A1)

We can recast the equation above in terms of mean densities and overdensities:

\[ n_\nu(1 + \delta_\nu(x)) = n_{\text{astro}}(1 + \delta_{\text{astro}}(x)) + n_{\text{atm}}(1 + \delta_{\text{atm}}(x)). \]

(A2)

By definition $\bar{n}_\nu = \bar{n}_{\text{astro}} + \bar{n}_{\text{atm}}$, and so

\[ \delta_\nu(x) = f_{\text{astro}} \delta_{\text{astro}}(x) + (1 - f_{\text{astro}}) \delta_{\text{atm}}(x), \]

(A3)

where $f_{\text{astro}} = \bar{n}_{\text{astro}}/\bar{n}_\nu$. Using Equation (2), we have

\[ \langle \delta_\nu(x) \delta_\nu(x') \rangle = \sum_\ell \frac{2 \ell + 1}{4 \pi} C_{\ell}^{\text{gg}} P_\ell(\cos \theta) \]

\[ \implies \langle \delta_\nu(x) f_{\text{astro}} \delta_{\text{astro}}(x') \rangle + (1 - f_{\text{astro}}) \delta_{\text{atm}}(x') = \sum_\ell \frac{2 \ell + 1}{4 \pi} \]

\[ \times (f_{\text{astro}} C_{\ell}^{\text{gg}} + (1 - f_{\text{astro}}) C_{\ell}^{\text{atm}}) P_\ell(\cos \theta). \]

(A4)

Comparing coefficients of $P_\ell(\cos \theta)$, we see that for a mixed population of astrophysical and atmospheric neutrino events with an $f_{\text{astro}}$ fraction of astrophysical neutrinos, the cross-correlation with the galaxy sample can be written in terms of the individual cross-correlations as

\[ C_{\ell}^{\text{gg}} = f_{\text{astro}} C_{\ell}^{\text{gg \ astro}} + (1 - f_{\text{astro}}) C_{\ell}^{\text{gg \ atm}}. \]

(A5)

Appendix B

IceCube Data Analysis and Synthetic Data Generation

B.1. IceCube All-sky Point-source Data in 2010–2012

The 2011–2012 data were taken with a full configuration of IceCube with 86 strings, while the 2010 data were taken with 79 strings. Each year of data contains about $10^5$ neutrino candidate events that passed the selection criteria described in Aartsen et al. (2017b). In the northern sky, which is defined by Aartsen et al. (2017b) as $\theta > 85^\circ$, the sample is mostly composed of neutrinos since up-going muons are shielded by Earth. It is dominated by atmospheric muons that follow a soft energy spectrum $dN/dE \propto E^{-3.7}$ (Honda et al. 2007; Aartsen et al. 2015b). In the southern sky, the public data are dominated by atmospheric muons with energies up to 10 PeV.

In each energy bin, we store the events in healpy maps with NSIDE = 128 in Celestial coordinates. The angular coordinates of a point on the sphere ($\theta, \phi$) are converted to the R.A. and decl. by $\theta = \pi/2 - \text{decl.}, \phi = \text{R.A.}$.

The counts map of the IceCube point-source data in the energy range $10^5$–$10^7$ GeV are shown in the left panel of Figure B1. The $\cos \theta$ distributions of the IceCube data in different energies are presented by the thick curves in the left panel of Figure B2. The first two energy bins are mostly composed of atmospheric neutrinos in the northern sky. The effective area in $[10^2, 10^3]$ GeV is much smaller than that in $[10^3, 10^4]$ GeV, and the first energy bin contains fewer events despite a greater atmospheric neutrino flux. Above 10 TeV, muons from the southern sky dominate the distribution. The three-year data do not have events above 10 PeV.

The autocorrelations of the overdensities of counts maps are shown in the right panel of Figure B2. The difference of the northern and southern skies, and declination-dependent differences in the same hemisphere lead to features at $\ell \lesssim 20$. The power spectra at $\ell \gtrsim 100$ are consistent with shot noise from Poisson statistics.

As the effective area of IceCube can vary notably across one decade of energy, we define a weighted effective area for a

\textsuperscript{10} https://github.com/KIPAC/muXgal
wide energy bin $i$ based on the effective area of finer bins $j$:

$$A_{\text{eff},i}(\cos \theta) = \frac{\sum_j A_{\text{eff},j}(\cos \theta)(E_{i,R}^{1-\alpha} - E_{i,L}^{1-\alpha})}{(E_{i,R}^{1-\alpha} - E_{i,L}^{1-\alpha})},$$  \hspace{1cm} (B1)

where $\alpha$ is the energy spectral index of events, and $E_{i,R}$ and $E_{i,L}$ are the energy at the right and left bounds of bin $j$. For comparison, the gray curve in the right panel of Figure B2 shows the power spectrum of the overdensity of the weighted effective area for IceCube between $10^4$ and $10^5$ GeV assuming events follow an $E^{-3.7}$ energy spectrum. The $C_\ell$ spectra do not depend strongly on the energy binning or spectral weighting.

### B.2. Generation of Synthetic Data

To generate synthetic data that contain astrophysical neutrinos, we first assume $f_\mu = 1$, and set the purity of astrophysical events in each energy bin, defined as

$$f_{\text{astro},i} \equiv \frac{N_{\text{astro},i}}{N_{\text{astro},i} + N_{\text{atm},i}}.$$  \hspace{1cm} (B2)

To be consistent with the muon neutrino population in the point-source public data, we adopt $(f_{\text{astro,1}}, f_{\text{astro,2}}, f_{\text{astro,3}}) = (2.2 \times 10^{-3}, 1.2 \times 10^{-2}, 0.15)$, which is based on the data and best-fit expectation from Monte Carlo simulation in the IceCube 10-year diffuse $\nu_\mu$ analysis (Figure 3 of Stettner 2019). Since we cannot differentiate the large muon population from neutrino events using the public data, we consider only the northern hemisphere for the astrophysical event generation and cross-correlation study in this work. The synthetic data are thus composed of $N_{\text{atm},i} = N_{\text{tot},i}(1 - f_{\text{astro},i})$ atmospheric events, which are generated the same way as pure atmospheric data, and $N_{\text{astro},i} = N_{\text{tot},i}f_{\text{astro},i}$ astrophysical events, as described below.

Assuming that astrophysical neutrino sources share the same sample variance as the galaxies, the probability of an astrophysical event from pixel $j$ is given by the density of the...
galaxies $\rho_j$, with $\sum_j \rho_j = 1$. To simulate detection of these astrophysical events, we consider each energy bin $i$ separately. For bin $i$, we evaluate the ratio of the effective area in each direction $j$ to the overall maximum effective area. We treat these relative probabilities of detection $p_{i,j}$ as absolute probabilities and generate a total of $N_{\text{astro},i} = \sum_j p_{i,j} \rho_j$ so that $N_{\text{astro},i}$ end up being detected.

The count map of a synthetic atmospheric event sample is compared with the actual IceCube data in Figure B1. The zenith distributions and power spectra are shown as thin curves in Figure B2. The difference in the thick and thin purple curves in the right panel of Figure B2 is due to the small event number in that energy bin.

Appendix C
Galaxy Catalogs

C.1. Analytical Power Spectrum

In order to generate an example of the expected $C_\ell$ of a sample of galaxies, we use the CLASS Boltzmann solver package.\textsuperscript{11} We choose the cosmological parameters to be the ones from the Planck 2018 best-fit cosmology (Planck Collaboration et al. 2018). We further assume that the sample of galaxies has a constant comoving number density, within a redshift range of $0.2 < z < 0.6$. Finally, we assume that the galaxy sample has a bias parameter of 1.2 with respect to the underlying matter fluctuations, and this value does not change as a function of redshift in the redshift range of interest. The analytical power spectrum is shown as a gray dashed curve in the right panel of Figure C1. A synthetic full-sky galaxy sample is drawn from the analytical power spectrum and used for testing of the method.

C.2. WISE–2MASS All-sky Infrared Galaxy Catalog

We combined the 2MASS color data with the WISE photometry data to improve efficiency of star–galaxy separation. We downloaded $\sim$5 million WISE–2MASS objects from the NASA/IPAC Infrared Science Archive website\textsuperscript{12} that satisfied the selection criteria suggested by Kovács & Szapudi (2015), and masked the region with Galactic latitude $|b| < 10^\circ$. The resulting galaxy sample is expected to have $<2\%$ stellar contamination and $>70\%$ galaxy completeness. The left panel of Figure C1 presents the counts map of the WISE–2MASS galaxy sample. The autocorrelation as computed with NSIDE = 128 is shown as the black curve in the right panel of the figure.

\textbf{Figure C1.} Left: number count map of the WISE–2MASS galaxy sample used in this work, constructed using the Wide-Field Infrared Survey Explorer (WISE: Wright et al. 2010) and the 2-Micron All-Sky Survey (2MASS; Skrutskie et al. 2006) infrared databases following Kovács & Szapudi (2015). The gray region denotes a mask of the Galactic plane with $|b| < 10^\circ$. The plot uses NSIDE = 128. Right: power spectra of the WISE–2MASS galaxy sample (black curve) and a synthetic galaxy sample based on matter fluctuations computed analytically. The difference in the power spectra is mainly due to the different redshift ranges of the galaxy samples. See Appendix B.1 for more details.

\textsuperscript{11} https://lesgourg.github.io/class_public/class.html

\textsuperscript{12} https://irsa.ipac.caltech.edu/Missions/wise.html
Appendix D
Projected 10-year Results

We present the projected results using synthetic 10-year point-source data in Figure D1. The synthetic data are generated using the cosine zenith distribution and effective area of the IceCube point-source data in 2012. Compared to the 3-year data, the 10-year data would contain more astrophysical events and better constrain $f_{\text{astro}}$. The median TS from $10^4$ realizations of synthetic data is $\sim 15$ in the optimistic scenario.

ORCID iDs
Ke Fang https://orcid.org/0000-0002-5387-8138
Eric Charles https://orcid.org/0000-0002-3925-7802
Yuuki Omori https://orcid.org/0000-0002-0963-7310

References
Aartsen, M. G., Abbasi, R., Abdou, Y., et al. 2013, Sci, 342, 1242856
Aartsen, M. G., Abbasi, R., Ackermann, M., et al. 2014, Jinst, 9, P03009
Aartsen, M. G., Abraham, K., Ackermann, M., et al. 2016, ApJ, 833, 3
Aartsen, M. G., Abraham, K., Ackermann, M., et al. 2017a, ApJ, 835, 45
Aartsen, M. G., Abraham, K., Ackermann, M., et al. 2017b, ApJ, 835, 151
Aartsen, M. G., Ackermann, M., Adams, J., et al. 2015a, arXiv:1508.05849
Aartsen, M. G., Ackermann, M., Adams, J., et al. 2015b, EPJC, 75, 116
Aartsen, M. G., Ackermann, M., Adams, J., et al. 2018, arXiv:1811.07979
Aartsen, M. G., Ackermann, M., Adams, J., et al. 2019a, arXiv:1911.11809
Aartsen, M. G., Ackermann, M., Adams, J., et al. 2019b, ApJ, 886, 12
Aartsen, M. G., Ackermann, M., Adams, J., et al. 2020a, PhRvL, 121, 051103
Aartsen, M. G., Ackermann, M., Adams, J., et al. 2020b, PhRvL, 124, 051103
Ahumada, R., Allende Prieto, C., Almíndez, A., et al. 2019, arXiv:1912.02905
Amann, M., Gruen, D., Regis, M., et al. 2020, PhRvL, 124, 101102
Ando, S., Tamborra, I., & Zandanel, F. 2015, PhRvL, 115, 221101
Bardeen, J. M., Bond, J. R., Kaiser, N., & Szalay, A. S. 1986, ApJ, 304, 15
Bechtol, K., Ahlers, M., Mauro, M. D., Ajello, M., & Vandenbroucke, J. 2017, ApJ, 836, 47
Bleem, L. E., Stalder, B., de Haan, T., et al. 2015, ApJS, 216, 27
Chang, C., Pujol, A., Mawdsley, B., et al. 2018, MNRAS, 475, 316
Clerc, N., Merloni, A., Zhang, Y.-Y., et al. 2016, MNRAS, 463, 4490
Fang, K., Kotera, K., Miller, M. C., Murase, K., & Oikonomou, F. 2016, JCAP, 2016, 017
Fang, K., Kotera, K., Murase, K., & Olinto, A. V. 2014, PhRvD, 90, 103005
Fang, K., & Murase, K. 2018, Nature, 559, 14
Feyereisen, M. R., Tamborra, I., & Ando, S. 2017, JCAP, 2017, 057
Górski, K. M., Hivon, E., Banday, A. J., et al. 2005, ApJ, 622, 759
Honda, M., Kajita, T., Kasahara, K., Midorikawa, S., & Sanuki, T. 2007, PhRvD, 75, 043006
Kaiser, N. 1984, ApJL, 284, L9
Kovács, A., & Szapudi, I. 2015, MNRAS, 448, 1305
Lenz, D., Doré, O., & Lagache, G. 2019, ApJ, 833, 75
López-Sanjuan, C., Tempel, E., Benítez, N., et al. 2017, A&A, 599, A62
Mertsch, P., Rameez, M., & Tamborra, I. 2017, JCAP, 2017, 011
Narayan, S. E., Sánchez, A. G., Prada, F., et al. 2013, MNRAS, 432, 743
Peebles, P. J. E. 1980, The Large-Scale Structure of the Universe (Princeton, NJ: Princeton Univ. Press)
Plack Collaboration, Aghamin, N., Akram, Y., et al. 2018, arXiv:1807.06209
Rieger, A. 2019, JCAP (Madison), 36, 1004
Skrutskie, M. F., Cutri, R. M., Stiening, R., et al. 2006, AJ, 131, 1163
Strittmatter, P. 2019, JCAP (Madison), 36, 1017
Tamborra, I., Ando, S., & Murase, K. 2014, JCAP, 2014, 043
Wilks, S. S. 1938, The Annals of Mathematical Statistics, 9, 60
Wright, E. L., Eisenhardt, P. R. M., Mainzer, A. K., et al. 2010, AJ, 140, 1868
Zonca, A., Singer, L., Lenz, D., et al. 2019, JOS, 4, 1298

Figure D1. Same as Figure 1 but for synthetic 10-year data. Left: posterior distributions of $f_{\text{astro}}$ found by a MCMC sampling of the parameter space. The blue lines mark the input values of $f_{\text{astro}}$. Right: the best-fit energy spectrum of astrophysical neutrinos that follow the galaxy sample used for the analysis.