

Aspects of spatially correlated random fields: Extreme-value statistics and clustering properties

Ka Hei Choi¹, James Creswell¹, Florian Kühnel^{1,2} and Dominik J. Schwarz³

¹*Arnold Sommerfeld Center, Ludwig-Maximilians-Universität,
Theresienstraße 37, 80333 München, Germany*

²*Max-Planck-Institut für Physik, Boltzmannstrasse 8, 85748 Garching, Germany*

³*Fakultät für Physik, Universität Bielefeld, Postfach 100131, 33501 Bielefeld, Germany*



(Received 3 April 2025; accepted 23 February 2026; published 10 March 2026)

Rare events of large-scale spatially correlated exponential random fields are studied. The influence of spatial correlations on clustering and nonsphericity is investigated. The size of the performed simulations permits to study beyond-7.5-sigma events (one in 10^{13}). As an application, this allows to resolve individual Hubble patches which fulfil the condition for primordial black hole formation. It is argued that their mass spectrum is drastically altered due to cocollapse of clustered overdensities as well as the mutual threshold-lowering through the latter. Furthermore, the corresponding nonsphericities may imply possibly large changes in the initial black hole spin distribution.

DOI: [10.1103/n3tr-t6dc](https://doi.org/10.1103/n3tr-t6dc)

I. INTRODUCTION

Spatially correlated random fields are fundamental to numerous physical phenomena, ranging from the microstructure of materials [1] to large-scale patterns in cosmology [2]. Understanding their statistical properties is important for describing for instance phase transitions [3], pattern formation [4] as well as other collective behaviors.

Classical extreme-value theory, pioneered by Fisher and Tippett [5] and formalized by Gnedenko [6] and Gumbel [7], primarily addresses independent and identically distributed random variables. However, many physical systems exhibit significant spatial correlations that invalidate the independence assumption [8]. These correlations can lead to substantial deviations from predictions of classical extreme-value statistics, affecting universality classes and scaling behaviors [9].

Clustering properties in random fields can be important in many cases, for instance near critical points where fluctuations become long-ranged [10]. Furthermore, the formation and distribution of clusters have profound implications for percolation processes [11] and transport phenomena in disordered media [12]. The interplay

between extreme events and clustering can, in general, give rise to emergent behaviors that are not evident when these aspects are considered independently (cf., Ref. [13]).

The statistics of peaks and clustering in Gaussian and near-Gaussian random fields have been investigated in the literature, especially with cosmological and astrophysical applications [14–18]. Our study extends works such as these by considering exponentially distributed discrete random fields. While for (spatially correlated) Gaussian random fields, not only many studies have been made, but, for a class of power spectra, even analytical formulas have been derived (cf., Ref. [19]), this is not the case for spatially correlated exponential random fields. The latter have become fundamentally important in cosmological contexts, especially for primordial black holes (PBHs) [20,21], whose probability distribution in many cases has exponential tails (cf., Refs. [22–24]). For our studies, we have generated data of a much larger size than any previous work, which enables further precision analysis of block maxima estimators for exponential random fields, for which no analytical expressions exist.

In this paper, we discuss results of our simulation of spatially correlated exponential random fields, being the currently largest one. We first focus on the extreme-value statistics emerging through block maxima, and then turn to the study of clustering in dependence of the strength of the spatial correlation. Here, we highlight the various emerging nonspherical shapes. This is then followed by a discussion of the application of our results to the formation of black holes in the early Universe. Finally, we conclude with a discussion of our results and an outlook.

Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI. Open access publication funded by the Max Planck Society.

II. SIMULATION OF CORRELATED RANDOM FIELDS

In this work, we have simulated more than 10^7 discrete random fields $F: \mathbb{R}^3 \rightarrow \mathbb{R}$, with each simulation having $N \times N \times N$ pixels for $N = 2^9$. The utilized class of power spectra is

$$P(k) \propto k^n, \quad (1)$$

where

$$\langle F_{\mathbf{k}} F_{\mathbf{p}}^* \rangle = \delta^3(\mathbf{k} + \mathbf{p}) P(k), \quad (2)$$

with $F_{\mathbf{k}}$ being the Fourier transform of $F(\mathbf{x})$, $k \equiv |\mathbf{k}|$; the spectral index n is taken to be either 0 (white noise), -1 , -2 or -3 , for definiteness. In order to address the infrared divergence of $P(k)$ for $n < 0$ at $k = 0$, we imposed an infrared cutoff to remove only the zero mode.

The generation of spatially correlated exponential random fields E is performed by first generating spatially correlated Gaussian random fields utilizing the discrete convolution method outlined in Ref. [25]: Let $\mathbf{m}$ denote an integer triplet with components $m_i \in [0, N-1]$ for $i = 1, 2, 3$, which specifies the grid positions $\mathbf{x}(\mathbf{m}) = (L/N)\mathbf{m}$ with a simulation cube of length L . Then

$$F(\mathbf{m}) = \frac{1}{N^3} \sum_{\boldsymbol{\kappa}} \exp\left(i \frac{2\pi}{N} \boldsymbol{\kappa} \cdot \mathbf{m}\right) \sqrt{P(k)} \xi(\mathbf{k}), \quad (3)$$

is a Gaussian random field of a given power spectrum $P(k)$, the dimensionless wave vector $\boldsymbol{\kappa} = \mathbf{k}L/(2\pi)$ has components $\kappa_i \in [-N/2, N/2]$ for $i = 1, 2, 3$, and $\xi(\mathbf{k}) = \exp[i\varphi(\mathbf{k})]$ is a stochastic field, with each randomized phase drawn independently from a uniform distribution, $\varphi(\mathbf{k}) \sim U[-\pi, \pi]$.

Following the generation of F , we proceed to map it to an exponential random field E in two steps. First, we map $F(\mathbf{m})$ into a uniform distribution $U(\mathbf{m}) \in [0, 1]$ by taking its cumulative distribution function (CDF). Second, we map $U(\mathbf{m})$ to an exponential random field $E(\mathbf{m})$ using the inverse CDF of the exponential distribution. This process defines the mapping $F(\mathbf{m}) \rightarrow E(\mathbf{m})$, with

$$E(\mathbf{m}) = -\frac{1}{\lambda} \ln \left\{ 1 - \frac{1}{2} \left[1 + \operatorname{erf} \left(\frac{F(\mathbf{m}) - \mu[F]}{\sqrt{2}\zeta[F]} \right) \right] \right\}, \quad (4)$$

where $\mu[F]$ and $\zeta[F]$ are the mean and standard deviation of F , respectively. This mapping aims to preserve the power spectrum and its spatial correlation, where the preservation of the power spectrum can be seen from the overlapping curves in Fig. 1. Note that, on one hand, this transformation is purely technical and is not intended to represent any physical process responsible for generating

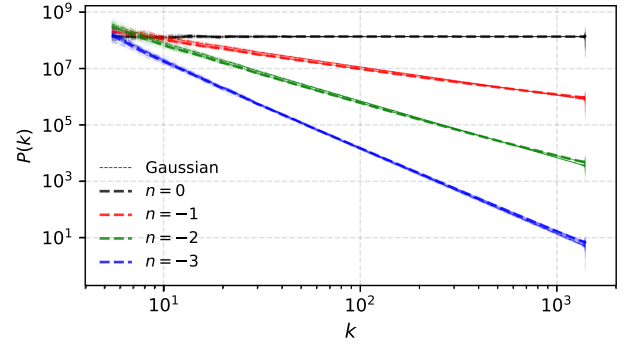


FIG. 1. Power-law power spectra given by Eq. (1) with spectral indices $n = 0, -1, -2, -3$ are shown for the Gaussian random field (dotted lines) and for the corresponding exponential random field (dash-dotted lines) obtained via the mapping in Eq. (4). For each value of n , we plotted 20 individual realizations of the Gaussian field and the power spectra of their corresponding mapped exponential random field. The less transparent curve for each n denotes the power spectrum averaged over realizations. Notice that, for all considered n , the Gaussian and exponential power spectra curves entirely overlap on top of each other, both at the level of individual realizations and therefore also in their average. This demonstrates that the transformation in Eq. (4) preserves the power-law spectrum for $n = 0, -1, -2, -3$ to a high accuracy.

non-Gaussianity. On the other hand, the exponential random field $E(\mathbf{m}) \in [0, \infty)$ should be considered as an effective model for studying the extreme value statistics of a random field that possesses an exponential tail rather than a complete perturbation distribution. In the following, for brevity, we will use $\mu = 0$ and $\zeta = \lambda = 1$ without loss of generality in the context of extreme value statistics, since our analysis concerns the probability distribution as a function of rarity, which is invariant under rescaling of λ .

III. EXTREME-VALUE STATISTICS

In many cases, one is interested in an average of a number of random variables. There are, however, many situations in which the relevant quantities are extreme fluctuations, such as *block maxima*

$$M_B \equiv \max_{i=1, \dots, B} \Delta_i, \quad (5)$$

with $B \in \mathbb{N}$ independently and identically distributed (iid) random variables $\{\Delta_i\}_{i=1}^B$, where in our work, the Δ_i correspond to the realizations of the random field F or E on the set of pixels within a given block [e.g., constituted by a number of Hubble patches in the case of PBH formation (see below)], sampled over a sufficiently large number of blocks. According to the Fisher-Tippett theorem [5], the distribution of M_B converges to one of the *generalized extreme value* (GEV) distributions, which are inherently non-Gaussian (see below) [26].

Specifically, if there exist two sequences $\{a_i > 0\}_{i=1}^{\infty}$ and $\{c_i \in \mathbb{R}\}_{i=1}^{\infty}$ such that

$$\lim_{i \rightarrow \infty} \text{Prob} \left[\frac{M_i - a_i}{c_i} < \delta \right] = H(\delta), \quad (6)$$

then $H(\delta)$ inevitably must be one of the GEV distributions (cf., Ref. [27])

$$-\log[H_{\alpha,\gamma}^s(\delta)] = \begin{cases} \frac{1}{\gamma} \left[1 + s \left(\frac{\delta - \alpha}{\gamma} \right) \right]^{-1/s} & (s \neq 0) \\ \frac{1}{\gamma} \exp \left[\left(\frac{\delta - \alpha}{\gamma} \right) \right] & (s = 0) \end{cases}. \quad (7)$$

The associated probability density function (PDF), $h_{\alpha,\gamma}^s$, is related to the CDF $H_{\alpha,\gamma}^s$ via

$$H_{\alpha,\gamma}^s(\delta) = \int_{-\infty}^{\delta} dy h_{\alpha,\gamma}^s(y). \quad (8)$$

Here, s , α , and γ are the *shape*, *location*, and *scale parameters*, respectively, to be determined from data. The choices $s = 0$, $s > 0$, and $s < 0$ correspond to the *Gumbel*, *Fréchet*, and *Weibull* distributions. Any distribution of iid random variables for which the limit exists, converging to one of those three distributions, belongs to their *maximum domain of attraction* (MDA) (cf. Ref. [27]). For $n = 0$, the Gaußian and the exponential distribution belong to the MDA of the Gumbel distribution.

Of course, for our studies with nonzero spectral index, the random variables are not truly independent, hence violating assumptions of the Fisher-Tippett theorem. However, as we will see below, the tails of the resulting block-maxima distributions will nevertheless follow a Gumbel distribution.

IV. CORRELATION LENGTH

When measuring the block maxima of spatially correlated random fields, the characteristic length scale of spatial correlation l_0 is necessary for selecting an appropriate box size. In this context, we refer to l_0 as the *correlation length*, being defined as (cf. Refs. [28,29])

$$l_0 \equiv \frac{\zeta_0}{\zeta_1}, \quad (9)$$

where ζ_0 and ζ_1 are the standard deviations of the random fields $F(\mathbf{x})$ or $E(\mathbf{x})$ and of their gradient, respectively. Intuitively, $l_0 = \langle k^2 \rangle_P^{-1/2}$ estimates spatial correlations by the weighted average of k^2 by $P(k)$ for continuous random fields. In numerical realizations, spatial derivatives are replaced by the central finite differences, such that $l_0 = \langle \sin^2(\mathbf{k} \cdot \Delta \mathbf{x}_i) / \Delta x_i^2 \rangle_P^{-1/2}$, for $i = 1, 2, 3$. This gives

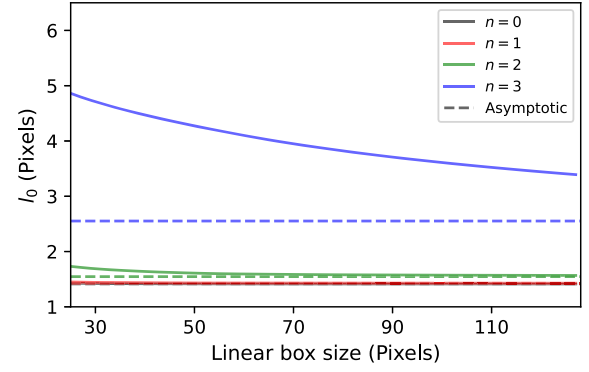


FIG. 2. Correlation length (9) of exponential random fields for different spectral indices, with a rare event of 6σ at the box centre. The flat dotted horizontal lines indicate the respective asymptotic correlation length derived from the full random field, while the solid lines depict the correlation length as a (diminishing) function of the box size. Note that the curves for $n = 0$ and $n = -1$ largely overlap.

$l_0 \sim \langle k^2 \rangle_P^{-1/2}$ for small k , while larger k (finer structures) are suppressed due to pixelization.

Figure 2 shows the numerical estimation of l_0 of exponential random fields with different spectral indices, guaranteed with a 6σ event at the box center. The flat line represents the asymptotic value, calculated using Eq. (9) by sampling over 10^4 exponential random fields. Decay of the curves with increasing block size can be observed, the smaller values of n yield a slowed decline.

In the following, we will consider a cubic block with linear size of 33 pixels. For $n = -3$, this is approximately 6 times the correlation length inside the block, and 12 times the asymptotic value (see Fig. 2). At this size, the curves for $n = 0, -1$ have already reached the asymptotic value, but not for $n = -2, -3$. This indicates that the chosen block

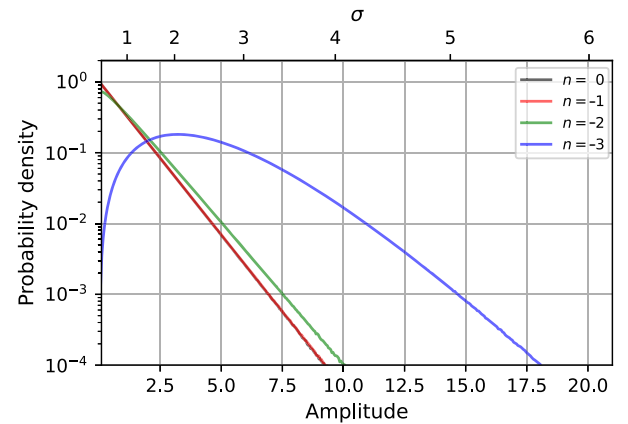


FIG. 3. Probability density distribution of exponential random fields *within* a block of 33 pixels (linear) dimension for different spectral indices (see legend). Note again the overlap of the curves for $n = 0$ and $n = -1$.

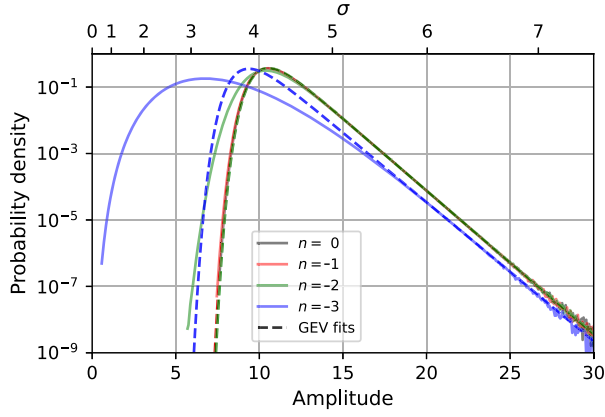


FIG. 4. Block-maxima PDF obtained by sampling exponential random fields of size $2^9 \times 2^9 \times 2^9$ pixels for each spectral index with rarity up to and beyond 7.5 sigma (1 in 10^{13}). The dashed line shows the corresponding best-fit GEV distribution [cf. Eq. (7)] between amplitude values 20 and 30. The parameters are $(\alpha, \gamma, s) = (10.52, 1.00, -9.84 \times 10^{-5} \simeq 0)[n = 0]$, $(10.51, 1.00, -3.27 \times 10^{-4} \simeq 0)[n = -1]$, $(10.54, 0.99, 9.86 \times 10^{-4} \simeq 0)[n = -2]$, and $(9.36, 1.04, 4.50 \times 10^{-4} \simeq 0)[n = -3]$. As before, the curves for $n = 0, -1$ essentially overlap.

size can capture substantial correlation for $n = -2, -3$, but is too large for $n = -1$ to detect any correlation effect.

The effect of spatial correlation is shown in Fig. 3, which shows the PDFs of the exponential random field within a block size of 33 pixels, and containing at least one extreme-value event with at least 6σ rarity. For $n = -3$, we observe a particularly strong suppression of low-amplitude

events in the neighborhood compared to the uncorrelated case ($n = 0$).

V. SPATIAL-CORRELATED RANDOM FIELDS

The PDF of block maxima is sensitive to spatial correlation, which is shown in Fig. 4. The ($n = 0$)-curve represents the PDF of block maxima sampled from an uncorrelated exponential field, which must follow a Gumbel distribution (cf., Ref. [27]). Any deviation from the case $n = 0$ in Fig. 4 is solely due to the effect of spatial correlation in the random field, as the spectral indices are the only varying parameter. The more the correlation length (within a block) deviates from its asymptotic value (cf., Fig. 2), the more the PDF of block maxima deviates from the case $n = 0$.

The key result of Fig. 4 is that spatial correlations tend to broaden the PDF of the block maxima, by enhancing the probability of finding low-amplitude block maxima while suppressing high-amplitude ones. This suppression occurs because spatial correlation causes rare events of extreme values to cluster, as can be observed in Fig. 3. However, block maxima only records the strongest signal within a given block, while the rest are discarded. This is in contrast with uncorrelated random fields ($n = 0$), where rare events are evenly spread, being hence more likely to be sampled as block maxima.

Similarly, for spatially correlated random fields, low-amplitude events tend to cluster and form large underdense regions, as can be well observed in Fig. 5, most pronounced for $n = -3$. Therefore, these underdense regions increase the likelihood of finding low-amplitude

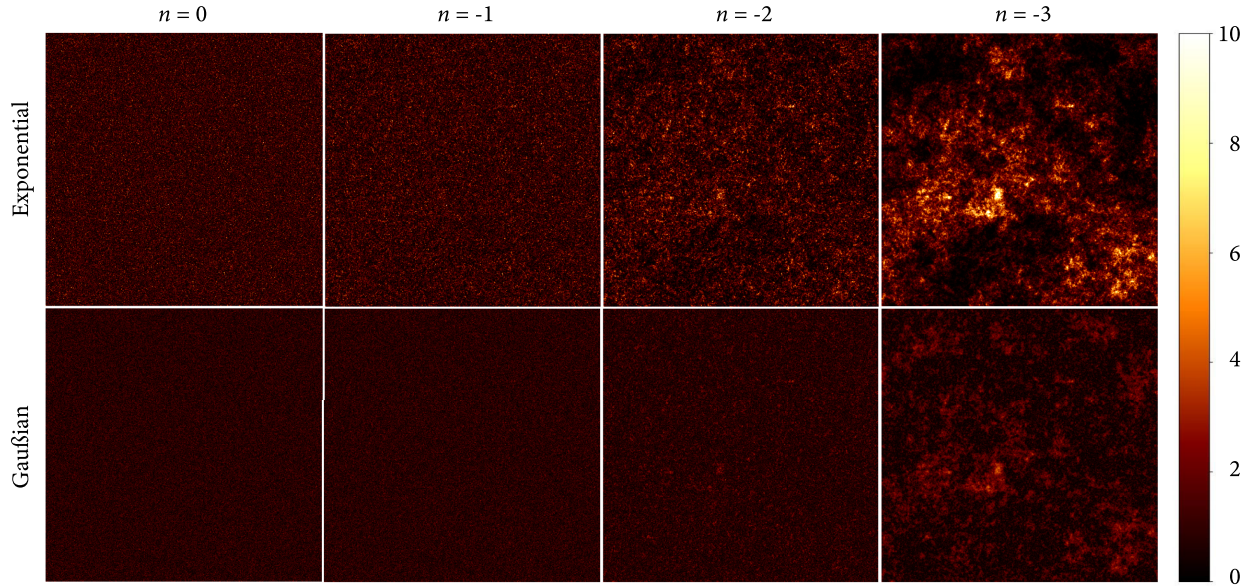


FIG. 5. Two-dimensional slices from a three-dimensional random field simulation, with the top row showing exponential random fields and the bottom row indicating Gaussian random fields, all cases have unit variance. The images, from left to right, display increasing spatial correlation in the random fields for spectral indices $n = 0, -1, -2, -3$, respectively. The color bar indicates the amplitude, with the brightest signals having a value of at least 10.

events as block-maxima. While spatial correlation broadens the PDF of block maxima, beyond the 6-sigma level, their tails show similar statistical properties across all spectral indices, exhibiting exponential decays at the same rate, shown as parallel slopes on the logarithmic scale in Fig. 4. This behavior arises because such extreme events are exceedingly rare, occurring at most once per simulation, or are rare enough to materialise only in one block. Accordingly, despite the presence of spatial correlation, rare events beyond the 6-sigma level are recorded as block maxima in a statistically independent manner and are well described by the tail of a Gumbel distribution ($s = 0$), as they are sampled independently from different exponential field simulation. This shows that α is the only parameter sensitive to changes in n for modeling tails behaviors in Fig. 4. Despite not being in the main focus of our studies, we note that also the block size influences the fitting parameters, with diminishing effect when the former increases.

VI. CLUSTERING

Spatial clustering of large peaks [30] generated by random fields plays important roles in many mathematical and physical contexts (cf., Ref. [16]). Depending on the specific type of the random field, in particular its spatial correlations, the clusters can emerge of various sizes and shapes. We will see that the spectral index [cf., Eq. (1)] plays an important role for these characteristics. Below we apply this to PBH formation and demonstrate the potential large impact of nonwhite-noise cases.

We have performed a cluster-shape analysis. For a pronounced demonstration of the resulting effects and to connect to subsequent analyses, we focus on clusters containing at least one 6-sigma signal in their centers. For our analysis, we will make use of the *weighted principal component analysis* (WPCA) [31], which allows us to quantify the sphericity in the random-field clusters in terms of distribution variance. A spherically distributed random field must exhibit equal variance along all principal components, whereas nonsphericity manifests as differences in them.

In contrast to the conventional principal component analysis (PCA) [32,33], which treats all data equally, WPCA discriminates individual data points and features by assigning weights to them. Specifically, WPCA minimises

$$\chi^2 = \sum_{i,j} W_{ij} \left(X_{ij} - \sum_k P_{ik} C_{kj} \right)^2. \quad (10)$$

Here, $X \equiv [x_1, \dots, x_B]$ is the grid matrix, with positions $x_i \equiv [x_i, y_i, z_i]^T$, where $i \in [0, B]$ for $i, B \in \mathbb{N}$, with B being the total number of grid points in the random-field block. For our purpose, each spatial coordinate must be weighted

equally by the field amplitude at a given position, such that $W \equiv [F(x_1)\mathbb{I}, \dots, F(x_B)\mathbb{I}] / \sum_i F(x_i)$, where $\mathbb{I} \equiv [1, 1, 1]^T$. The quantity χ^2 is then minimized by fitting the coefficient matrix C with the principal-components matrix P , whose rows and columns correspond to spatial coordinates and components, respectively; the rows of C correspond to components and its columns to grid points.

The principal components can be diagonalized into the distribution-variance matrix $\Sigma \equiv CC^T$, with $\Sigma_i \geq \Sigma_{i+1} \forall i \in \mathbb{N}$. These variances measure the sphericity of the random field within a block by measuring its anisotropy. The ratio $\Sigma_{31} \equiv \Sigma_3 / \Sigma_1 \in [0, 1]$ reflects isotropy when $\Sigma_{31} = 1$, and anisotropy for $\Sigma_{31} < 1$. If clusters form due to spatial correlation, Σ_{31} indicates whether the cluster is spherically symmetric. In the presence of anisotropy, the cluster shape can be characterized by the ratio $\Sigma_{32} \equiv \Sigma_3 / \Sigma_2 \in [0, 1]$: $\Sigma_{32} = 1$ implies an ellipsoidally shaped cluster, $\Sigma_{32} \sim \Sigma_{31}$ (black line) indicates a circular disk shape; $\Sigma_{32} < 1$ corresponds to a disklike shape.

Figure 6 illustrates the ratios of principal components characterizing the geometric properties of the random field. It is clear that nonsphericity develops as spatial correlation increases. In the absence of spatial correlation ($n = 0$), both Σ_{31} and Σ_{32} have mean values closer to 1 of minor standard deviation, indicating the random field is mostly homogeneous and isotropic. As the spectral index increases, particularly at $n = -3$, the mean of Σ_{31} and Σ_{32} decrease, indicating that the formed cluster is preferably anisotropic. Furthermore, as the standard deviation increases, it indicates a large variety in the cluster shapes, either forming elongated or flattened shapes.

To isolate the effect of the tail contribution from that of the spatial correlations, we have also performed simulations of Gaussian random fields with the same set of power spectra as for exponential fields (see the two panels of Fig. 6 for a direct comparison). Our results show that non-Gaussianity amplifies the degree of nonsphericity generated by spatial correlation, and produces a broader variety of cluster shapes. For $n = -2$ and $n = -3$, where large spatial clusters are formed, the exponential tails in the exponential random field enhance the degree of spatial anisotropies, even though these Gaussian and exponential random fields shares identical power spectra by construction of Eq. (4). This occurs because the nonsphericity measure (10) weights the amplitudes of the random field within each block. Hence, in the exponential case, the existence of a heavy tail leads to a greater variance of amplitude within a cluster; therefore, the presence of an at least 6σ event within a cluster can change Σ_{ii} , for $i = 1, 2, 3$, more substantially when compared to the Gaussian case.

VII. APPLICATION: PRIMORDIAL BLACK HOLES

Primordial black holes [20,21] form in the early Universe, usually before matter-radiation equality and have

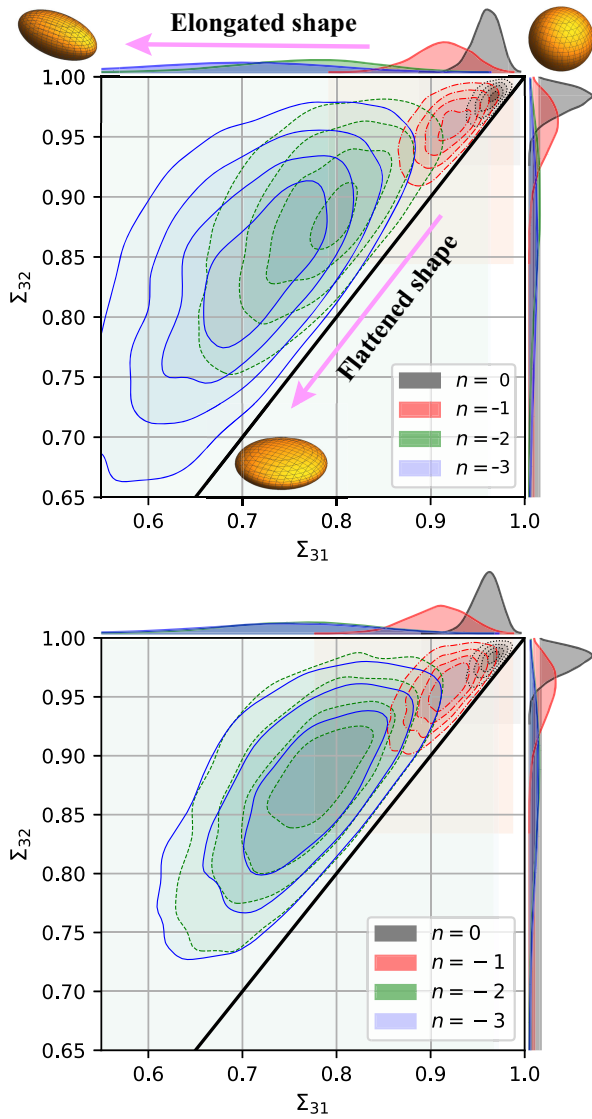


FIG. 6. Ratio of weighted principal components Σ_{32} against Σ_{31} obtained from sampling at least 10^5 simulations that contain at least one extreme-value event of 6 sigma rarity. The black straight line marks $\Sigma_{31} = \Sigma_{32}$. Equal probability contours are shown, representing the weighted principal components for a cluster with rare fluctuation (of 6 sigma and beyond) in its center. The upper panel shows the results for exponential random fields, while the lower panel depicts the Gaussian case.

long been discussed as dark matter candidates [34]; see Ref. [35] for a recent review. After the detection of gravitational waves from black hole binaries [36], it was realized these could be primordial in Nature [37].

There are numerous formation scenarios (see Ref. [38] for the currently most extensive review), but all have in common that they yield overdensities which are sufficiently large to collapse to a black hole. This could have happened for Hubble-horizon-sized overdensities with density contrast δ exceeding a certain critical threshold δ_c , which depends on the shape of the overdensities and their

statistics, the collapsing medium (with particular sensitivity to the equation of state [39]) as well as the presence of nearby overdensities [40]. For an isolated density peak, assuming spherical symmetry and collapse during radiation domination, simulations give $\delta_c = 0.45$ [41,42]. Since this value is about four orders of magnitude larger than that of typical CMB fluctuations [44], PBH formation fundamentally depends on the tail of the density distribution, making it inherently rare, usually beyond five sigma. Importantly, the collapse is not instantaneous, typically taking several e-folds (see Refs. [41,45,46]). Especially if other large nearby overdensities are present, it can take around 10^3 Hubble times until the final black hole mass is reached [40].

Here, we will study exactly the effect of finite formation time. This implies the consideration of blocks constituted by the multiple initial Hubble volumina (from the time of horizon crossing) contained in the Hubble patch at the later time of PBH formation. The nature of gravitational collapse implies that the largest of the involved overdensities will be key for the final black hole. Hence, this situation naturally leads to considering block maxima. It has been shown that the probability distributions in many cases have exponential tails (cf., Refs. [22–24]), these being in our main focus.

At present, there exists no measurement of the primordial power spectrum on scales k_{PBH} relevant for PBH formation. It is clear, however, that the near-scale-invariant form determined at CMB scales cannot hold around k_{PBH} , but must rather dramatically increase to yield enough sufficiently large overdensities undergoing gravitational collapse. Due to the mentioned uncertainty, one has to make assumptions, and here we will assume that for a range of scales relevant for PBH formation, the primordial power spectrum can be approximated by the power-law form of Eq. (1). This is somewhat arbitrary, but our results will qualitatively hold whenever the slope of the power spectrum is similar.

Due to the partly strong clustering, especially in the case of $n = -3$, it is important to consider the effect of neighboring overdensities on the PBH formation. Conveniently, Ref. [40] has recently performed a detailed numerical study on how the presence of an additional overdensity alters the black hole formation process. For our analysis, we extract the final PBH mass, as a function of horizon mass, from the numerical results in Table 1 of the mentioned paper. Together with its Fig. 9, and subsequent interpolation, we obtain an approximation of the PBH mass function resulting from two overlapping overdensities. At present, no extension of the work [40] exists to go beyond two overlapping overdensities, which would be needed in our strongly clustered cases. We will hence naïvely assume that, to lowest order, the individual effects of each additional overdensity are additive, knowing that such situations can only be properly addressed through (yet to be performed) numerical simulations. Hence, our results—despite being qualitatively correct—should quantitatively

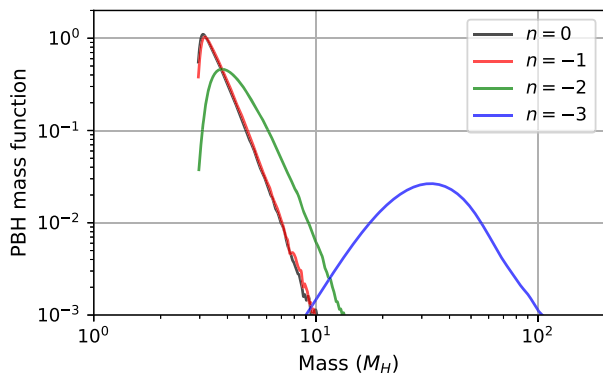


FIG. 7. Probability density distribution of the PBH mass (in unit of horizon mass M_H) obtained from sampling at least 10^5 exponential field simulations that contain at least one extreme-value event of 6 sigma rarity. The mass spectrum considers gravitational collapse of the beyond-6-sigma fluctuations, taking neighboring overdensities into account.

be taken *cum grano salis*. On the one hand, the assumed additivity may overestimate our results; on the other, we have conservatively neglected two effects which clearly underestimate the strength of our results: first, the presence of non-nearest neighbors, and, second, the accretion phase which Escrivà-Yoo [40] show to significantly increase the black hole mass, this being possibly enhanced by the clustering of overdensities. A more precise evaluation is needed for a full quantitative understanding of this new effect.

Figure 7 shows our findings for the PBH mass functions—as before, for the four cases of white noise ($n = 0$) and nonzero spatial correlations ($n = -1, -2, -3$). As mentioned, as a rough approximation, we have assumed that the effects on threshold lowering and collapse-time prolongation, leading to an overlap-induced increase of the PBH mass, are additive for additional neighboring overdensities. It can be observed that the latter can have a large effect in various ways, yielding a shift, broadening [47] and lowering of the PBH mass function whose minimum mass increases with increasing spatial correlation. This is due to the latter enabling the process of absorption of nearest neighboring peaks. Of course, the exact numerical values remain to be obtained from a detailed simulation, but one can be certain to expect the effect to remain and to be at least of similar magnitude. There are, of course, several other effects which alter the shape of the mass function, such as a softening of the equation of state (cf., Ref. [39]). While most of these alter the number of PBHs at a certain mass, the novel effect described in this work additionally implies a substantial shift toward larger PBH masses [48].

Depending on the value of the spectral index, even two above-threshold peaks can occur right next to each other. Concretely, for six-sigma fluctuations, the corresponding probabilities are 10^{-9} for $n = 0$, 10^{-7} for $n = -1$, 10^{-4} for $n = -2$, and 10^{-2} for $n = -3$. The associated directly

neighboring overdensities will clearly lead to the formation of a single black hole instead of separate ones. Similarly, also slightly more distant overdensities can be expected to coalesce. Besides nearest neighbors, in a strongly clustered situation, also more distant high-amplitude fluctuations can be assumed to be relevant. In general, the cluster-induced nonsphericities (cf., Fig. 6) can be expected to substantially alter the initial PBH spin distribution, leading to larger initial spins than previously expected (cf., Ref. [49] for prior work studying the spin distribution of clustered PBH populations).

Furthermore, a fraction of separate, but sufficiently nearby, PBHs can be expected to undergo merger processes, thereby implying additional changes in spin and mass. These effects will be left for future work, which will also include the study of more realistic primordial power spectra. Indeed, PBH formation without overproduction cannot originate from pure power-law power spectra studied in this work, but must involve one (or several) peak(s). Any peak has a rising and a falling side, for which power-law power spectra may serve as a first approximation. While this might be very rough, the purpose of our work is to show as a proof of principle that an important class of power spectra will have a strong effect on clustering and sphericities as well as the PBH mass function. These novel effects will certainly be present for any peaked primordial power spectrum which leads to PBH formation. Finally, we would like to point out that, since all studied effects predominantly involve overdensities deep in the distribution of the tail, they would all be affected similarly by an overall normalization.

VIII. CONCLUSION

In this paper, we have studied large-scale spatially correlated exponential random fields, particularly investigating the rare-event statistics of clustering and nonsphericity developed by the spatial correlation and the existence of an exponential tail. For our studies, we performed the currently largest corresponding simulation, permitting to resolve beyond-7.5-sigma events (1 in 10^{13}). In the context of extreme-value theory, we have shown that rare events in spatially correlated exponential random fields can be modeled using generalized extreme-value distributions.

We have further demonstrated that, for sufficiently rare events, this distribution reduces to the same family of extreme-value distributions as in their independent case for rare events, i.e., the Gumbel distribution. More generally, if a sequence of random events is large compared to the correlation scale of the random variable, then it is possible to extend the application of Gumbel distribution to describe random variables that are identical but not independently distributed even in nonasymptotic models.

Therefore, as an application, we were able to study individual Hubble patches which would fulfill the

condition for the formation of primordial black holes. Here, we point out a novel effect manifesting in sizeable spatial-correlation-induced shift, broadening and lowering of their mass spectrum. The emerging nonsphericity will likely also alter their initial spin distribution, which we will investigate in a forthcoming publication. Extending the applications of extreme-value distributions to other cosmological situations in which spatially correlated random fields are relevant may reveal other novel effects and will provide further insights into the statistical nature of rare events in our Universe.

ACKNOWLEDGMENTS

We thank the anonymous referee for useful input. D. J. S. acknowledges support of Deutsche Forschungsgemein-

schaft (DFG, German Research Foundation)—Project No. 315477589–TRR 211. K. H. C. acknowledges the supports of the Excellence Cluster ORIGINS. The authors gratefully acknowledge the *Gauß Centre for Supercomputing e.V.* [50] for funding this project by providing computing time on the *GCS Supercomputer SuperMUC at Leibniz Supercomputing Centre* [51].

DATA AVAILABILITY

The data that support the findings of this article are not publicly available upon publication because it is not technically feasible and/or the cost of preparing, depositing, and hosting the data would be prohibitive within the terms of this research project. The data are available from the authors upon reasonable request.

-
- [1] S. Torquato, *Random Heterogeneous Materials: Microstructure and Macroscopic Properties* (Springer, New York, 2002).
 - [2] P. J. E. Peebles, *The Large-Scale Structure of the Universe* (Princeton University Press, Princeton, NJ, 1980).
 - [3] H. E. Stanley, *Introduction to Phase Transitions and Critical Phenomena* (Oxford University Press, New York, 1971).
 - [4] M. C. Cross and P. C. Hohenberg, *Rev. Mod. Phys.* **65**, 851 (1993).
 - [5] R. A. Fisher and L. H. C. Tippett, *Math. Proc. Cambridge Philos. Soc.* **24**, 180 (1928).
 - [6] B. V. Gnedenko, *Ann. Math.* **44**, 423 (1943).
 - [7] E. J. Gumbel, *Statistics of Extremes* (Columbia University Press, New York, 1958).
 - [8] J.-P. Bouchaud and M. Mézard, *J. Phys. A* **30**, 7997 (1997).
 - [9] S. T. Bramwell, P. C. W. Holdsworth, and J.-F. Pinton, *Nature (London)* **396**, 552 (1998).
 - [10] D. Stauffer and A. Aharony, *Introduction to Percolation Theory*, 2nd ed. (Taylor & Francis, London, 1994).
 - [11] A. Coniglio, *J. Phys. A* **15**, 3829 (1982).
 - [12] S. Havlin and D. Ben-Avraham, *Adv. Phys.* **51**, 187 (2002).
 - [13] D. Sornette, *Critical Phenomena in Natural Sciences: Chaos, Fractals, Self-Organization and Disorder: Concepts and Tools* (Springer, New York, 2000).
 - [14] J. A. Peacock and A. F. Heavens, *Mon. Not. R. Astron. Soc.* **217**, 805 (1985).
 - [15] Y. Hoffman and J. Shaham, *Astrophys. J.* **297**, 16 (1985).
 - [16] J. M. Bardeen, J. R. Bond, N. Kaiser, and A. S. Szalay, *Astrophys. J.* **304**, 15 (1986).
 - [17] P. Coles, *Mon. Not. R. Astron. Soc.* **238**, 319 (1989).
 - [18] T. Matsubara, *Phys. Rev. D* **101**, 043532 (2020).
 - [19] R. J. Adler, *The Geometry of Random Fields* (Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, USA, 2010).
 - [20] Y. Zel'dovich and I. Novikov, *Sov. Astron.* **10**, 602 (1967), <https://ui.adsabs.harvard.edu/abs/1967SvA....10..602Z/abstract>.
 - [21] B. J. Carr and S. W. Hawking, *Mon. Not. R. Astron. Soc.* **168**, 399 (1974).
 - [22] M. Biagetti, G. Franciolini, A. Kehagias, and A. Riotto, *J. Cosmol. Astropart. Phys.* **07** (2018) 032.
 - [23] J. M. Ezquiaga, J. García-Bellido, and V. Vennin, *J. Cosmol. Astropart. Phys.* **03** (2020) 029.
 - [24] Y. Tada and V. Vennin, *J. Cosmol. Astropart. Phys.* **02** (2022) 021.
 - [25] E. Bertschinger, *Astrophys. J. Suppl. Ser.* **137**, 1 (2001).
 - [26] It is important to stress the nonperturbative nature of the non-Gaussianity in this work. In cosmology, the commonly used parameters like f_{NL} or g_{NL} describe an expansion around Gaussianity. While it might well be that for a given rarity, such as 6σ , the mentioned expansion might actually be sufficient for the derivation of the studied effects, the exact exponential tail, however, cannot be obtained from any finite expansion based on Gaussianity.
 - [27] P. Embrechts, T. Mikosch, and C. Klüppelberg, *Modelling Extremal Events: For Insurance and Finance* (Springer-Verlag, Berlin, Heidelberg, 1997).
 - [28] P. D. Naselsky and D. I. Novikov, *Astrophys. J.* **507**, 31 (1998).
 - [29] J. Kasak, J. Creswell, H. Liu, and P. Naselsky, *Phys. Rev. D* **104**, 023502 (2021).
 - [30] We appreciate that most of the literature discusses continuous situations in which a *peak* is an extended feature (unless a delta-peak is assumed). On the contrary, in our discrete study we refer to a peak as an excess in a single pixel (representing an initial Hubble patch at the time of horizon crossing of an overdensity).
 - [31] L. Delchambre, *Mon. Not. R. Astron. Soc.* **446**, 3545 (2014).
 - [32] K. Pearson, *Philos. Mag.* **2**, 559 (1901).
 - [33] H. Hotelling, *J. Educ. Psychol.* **24**, 417 (1933).
 - [34] G. Chapline, *Nature (London)* **253**, 251 (1975).
 - [35] B. Carr and F. Kuhnel, *Annu. Rev. Nucl. Part. Sci.* **70**, 355 (2020).

- [36] B. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), *Phys. Rev. X* **6**, 041015 (2016); **8**, 039903(E) (2018).
- [37] S. Bird, I. Cholis, J. Muñoz, Y. Ali-Haïmoud, M. Kamionkowski, E. D. Kovetz, A. Raccanelli, and A. G. Riess, *Phys. Rev. Lett.* **116**, 201301 (2016).
- [38] A. Escrivá, F. Kühnel, and Y. Tada, in *Black Holes in the Era of Gravitational-Wave Astronomy* (Elsevier, New York, 2024), pp. 261–377.
- [39] B. Carr, S. Clesse, J. García-Bellido, and F. Kühnel, *Phys. Dark Universe* **31**, 100755 (2021).
- [40] A. Escrivà and C.-M. Yoo, *J. Cosmol. Astropart. Phys.* **04** (2024) 048.
- [41] I. Musco and J. C. Miller, *Classical Quantum Gravity* **30**, 145009 (2013).
- [42] More advanced estimates utilize the so-called compaction function (see, e.g., Ref. [43])—an approach which unavoidably induces some degree of smoothing, hence—given the discrete (pixelated) approach used in this work—it is a relevant question to ask whether and how the use of the smoothing due to the use of compaction functions alters the results. This could certainly have large implications on subhorizon scales, but in our work, each pixel is thought to represent an entire initial Hubble patch, so we do not expect a strong effect.
- [43] M. Shibata and M. Sasaki, *Phys. Rev. D* **60**, 084002 (1999).
- [44] Y. Akrami *et al.* (Planck Collaboration), *Astron. Astrophys.* **641**, A10 (2020).
- [45] I. Musco, J. C. Miller, and A. G. Polnarev, *Classical Quantum Gravity* **26**, 235001 (2009).
- [46] A. Escrivà Mañas, Numerical simulations of primordial black holes, Ph.D. thesis, University of Barcelona, 2021.
- [47] Observe that the broadening of the PBH mass function does not only extend toward higher masses, due to the absorption of clustered peaks, but also toward lower masses, since neighboring (partly overlapping) peaks generically lower the PBH formation threshold (see Ref. [40]).
- [48] Note that this effect cannot directly compare to the Gaussian random field case, since Gaussian random fields do not possess a heavy tail. Specifically, if the threshold for PBH formation is set to 20 standard deviation (6σ rarity) of the exponential random field, then a Gaussian random field of the same variance would yield essentially no threshold exceedance.
- [49] V. De Luca, V. Desjacques, G. Franciolini, and A. Riotto, *J. Cosmol. Astropart. Phys.* **11** (2020) 028.
- [50] www.gauss-centre.eu.
- [51] www.lrz.de.