

Fast generation of Gaussian random fields for direct numerical simulations of stochastic transport

D. I. Palade* and M. Vlad†

*National Institute of Laser, Plasma and Radiation Physics,
PO Box MG 36, RO-077125 Măgurele, Bucharest, Romania*

Abstract

We propose a novel discrete method of constructing Gaussian Random Fields (GRF) based on a combination of modified spectral representations, Fourier and Blob. The method is intended for Direct Numerical Simulations of the V-Langevin equations. The latter are stereotypical descriptions of anomalous stochastic transport in various physical systems. From an Eulerian perspective, our method is designed to exhibit improved convergence rates. From a Lagrangian perspective, our method offers a pertinent description of particle trajectories in turbulent velocity fields: the exact Lagrangian invariant laws are well reproduced. From a computational perspective, our method is twice as fast as standard numerical representations.

* dragos.palade@infpr.ro

† madalina.vlad@infpr.ro

I. INTRODUCTION

Stochastic phenomena are ubiquitous in nature and laboratory, being present in various sciences: physics and chemistry [1], biology [2], finances [3], social sciences [4], etc. In particular, physical stochastic processes such as turbulent flows [5], anomalous transport in fusion plasmas [6, 7], flows through porous media [8], seismic motion [9] are complex phenomena that are modeled by nonlinear stochastic (partial) differential equations [10]. Most of the theoretical studies [11–14] are based on direct numerical simulations (DNSs) (or Monte Carlo simulations). Unfortunately, the ensemble statistics for the input processes as well as for the solutions exhibit slow convergence rates, with fluctuations that decay, usually, as $M^{-1/2}$, where M is the dimension of the ensemble (the number of realizations). Thus, the numerical effort involved in a DNS is a matter of concern, even in the context of the computing power available nowadays [15].

The Gaussian random fields (GRFs) [16] are input stochastic processes for a large class of models. A DNS requires the generation of a large ensemble of fields, which represents an important fraction of the computing time. The topic of GRFs representation is old and a large amount of constructing techniques are available [9, 17–19]. From all those, by far, the most employed (especially in the context of DNSs of stochastic transport and trajectory statistics [6, 10, 11, 20, 21]) is the spectral method of discrete Fourier decomposition [9] implemented using the fast Fourier transform (FFT) algorithms. This standard method exhibits some important flaws. Mathematically, the resulting fields and their covariance functions are periodic, a feature that is not usually required by the statistical model. Computationally, the convergence of the statistical properties is slow. Physically, invariants of motion may be altered due to in-between grid points interpolation.

The aim of the present study is to derive and to analyze methods of generation of GRFs having as main criterion the minimization of the computation time. We propose a novel method (the hybrid Fourier-Blob (FB) representation) that has several advantages compared to the FFT representation. It strongly improves the convergence rates of the field statistics without imposing periodicity. These improvements reduce the computation time compared to the FFT method by as much as an order of magnitude.

The theoretical results are presented in Section II. Starting from general, integral, representations of a GRF, we derive discrete variants. They are modified by introducing additional

random elements, which are expected to improve the convergence. Two particular types of representations are chosen (Fourier- and Blob-like), which are shown to be canonically conjugated.

Section III contains a detailed study of the accuracy of these discrete representations. Three variants of the Fourier and Blob representation are considered. Their ability to reproduce the characteristics of the GRFs is numerically analyzed at two levels. The basic level consists of a comparison of the results on the standard Eulerian quantities (covariance and distribution functions). The second level involves the DNS of a special test-particle advection process: particle stochastic motion in two-dimensional, incompressible, time-independent velocity fields. This is a Hamiltonian process with two Lagrangian invariants. One appears in each trajectory (local invariant) and the other involves the distribution of the Lagrangian velocity (statistical invariant). They provide strong benchmarks for the numerical simulation and, implicitly, for the GRF generation method. The results on the diffusion coefficients and on the distribution of the displacements are also compared. Finally, this detailed analysis permitted to find the hybrid FB method, which appears as the fastest GRF generator able to be used in DNS studies of complex stochastic advection processes. The conclusions are summarized in Section IV.

II. THEORY

We consider a real GRF $\phi(\mathbf{x})$ on a d dimensional space $\phi : \mathbb{R}^d \rightarrow \mathbb{R}$ with zero average $\langle \phi(\mathbf{x}) \rangle = 0$ and covariance function $\mathcal{E}(\mathbf{x}; \mathbf{y}) = \langle \phi(\mathbf{x})\phi(\mathbf{y}) \rangle$, where $\langle \cdot \rangle$ is the statistical averaging operation. This field can be generally represented through a set of parametric functions $\{F(\mathbf{x}; \mathbf{s})\}$ as:

$$\phi(\mathbf{x}) = \int d\mathbf{s} F(\mathbf{x}; \mathbf{s}) \zeta(\mathbf{s}) \quad (1)$$

$$\mathcal{E}(\mathbf{x}; \mathbf{y}) = \int d\mathbf{s} F(\mathbf{x}; \mathbf{s}) F(\mathbf{y}; \mathbf{s}) \quad (2)$$

where $\zeta(\mathbf{s})$ is an uncorrelated random variable $\langle \zeta(\mathbf{s})\zeta(\mathbf{s}') \rangle = \delta(\mathbf{s} - \mathbf{s}')$. It can be easily proven that Eqs. (1), (2) reproduce the correct covariance function $\mathcal{E}(\mathbf{x}; \mathbf{y})$ while its Gaussian character is guaranteed by the Central Limit Theorem. Before discussing the nature of the parametric functions $F(\mathbf{x}; \mathbf{s})$ let us address the matter of discreteness.

It is tempting to pass from the integral representation (1), (2) to a finite and discrete form in two steps: truncate and discretize the domain of integration. As we shall see, the operator $F(\mathbf{x}; \mathbf{s})$ can be, usually, safely neglected outside some finite domain in the \mathbf{s} space so the truncation is justified. But using a Riemann sum $\int d\mathbf{s} \rightarrow \sum_{\mathbf{s}}$ to approximate the integrals in Eqs. (1), (2) might not be the best approach (in the sense of errors, smoothness and convergence).

A. Discrete representations

Let us define in the \mathbb{R}^d parametric space $\{\mathbf{s}\}$ an equidistant grid of points $\{\mathbf{s}_0\}$ with the interspacing L , such that each point \mathbf{s}_0 is centered in the hypercubic domain $\mathcal{D}(\mathbf{s}_0)$ of volume L^d . Accordingly, the integral over parameters can be broken as:

$$\int d\mathbf{s} \equiv \sum_{\mathbf{s}_0} \int_{\mathcal{D}(\mathbf{s}_0)} d\mathbf{s}.$$

Considering that $F(\mathbf{x}; \mathbf{s})$ are infinitely differentiable, one can Taylor expand around a grid point \mathbf{s}_0

$$F(\mathbf{x}; \mathbf{s}) = \sum_n \frac{(\mathbf{s} - \mathbf{s}_0)^n}{n!} \nabla_{\mathbf{s}_0}^n F(\mathbf{x}; \mathbf{s}_0),$$

and the field (1) can be written

$$\phi(\mathbf{x}) = \sum_{\mathbf{s}_0} \left(\sum_n \frac{\hat{\alpha}_n}{n!} \nabla_{\mathbf{s}_0}^n \right) F(\mathbf{x}; \mathbf{s}_0),$$

where the coefficients

$$\hat{\alpha}_n = \int_{\mathcal{D}(\mathbf{0})} d\mathbf{s} \zeta(\mathbf{s}) \mathbf{s}^n$$

are random with zero average and correlation $\langle \hat{\alpha}_n \hat{\alpha}_m \rangle \propto L^{2d} (L/2)^{n+m+1} / (n+m+1)$. In essence, we pass from an integral (dense) representation (1) to a discrete one by recasting the dense character in an infinite series of random variables $\hat{\alpha}_n$. In the limit $L \rightarrow 0$ we can cut the series at the first order $\mathcal{O}(L^2)$ and approximate:

$$\phi(\mathbf{x}) \approx \sum_{\mathbf{s}_0} \alpha_0 F(\mathbf{x}; \mathbf{s}_0 + \beta) \tag{3}$$

where α_0 is a Gaussian variable with L^d variance while β a Cauchy distributed variable with the scale parameter $L/\sqrt{12}$. The representation (3) reproduces the correlation only in an approximate manner (dependent on the magnitude of L). Also, the density of points

$$\rho = \sum_{\mathbf{s}_0} \langle \delta[\mathbf{x} - (\mathbf{s}_0 + \beta)] \rangle \quad (4)$$

is a periodic fluctuating profile around the average $\rho = 1/L^d$ with roughly $\rho/4$ amplitude.

We propose a representation of the GRF that has the structure of (3) and eliminates the above disadvantages:

$$\phi(\mathbf{x}) \approx L^{d/2} \sum_j \zeta_j F(\mathbf{x}; \mathbf{s}_j) \quad (5)$$

where the random variables ζ_j are uncorrelated $\langle \zeta_j \zeta_i \rangle = \delta_{i,j}$ and the points \mathbf{s}_j are uniformly random distributed with the average density $\rho = 1/L^d$.

B. Gaussian convergence

The discrete form (5) reproduces the exact covariance function even in the limit $L \rightarrow \infty$ (a single term in the sum). The limit $L \rightarrow 0$ (large $\rho = 1/L^d$) is required in order to achieve the multivariate Gaussian probability distribution function $P(\{\varphi_i\}; \{\mathbf{x}_i\}) = \langle \prod_{i=1}^k \delta[\varphi_i - \phi(\mathbf{x}_i)] \rangle$. Our aim is to maximize the convergence rate towards Gaussian character of the series (5) at a fixed density. In other words, we look for representations of $\phi(\mathbf{x})$ that are "Gaussian enough" with a minimal parametric density ρ .

We focus for simplicity further on the one-point PDF $P(\varphi; \mathbf{x}) = \langle \delta[\varphi - \phi(\mathbf{x})] \rangle$. The local rate of convergence for a sum of independent variables is bounded by the Berry-Esseen theorem [22] as $r < C\sigma_3/\sigma_2$. In our case:

$$\begin{aligned} \sigma_2^2 &= \sum_j \langle |\zeta_j F(\mathbf{x}; \mathbf{s}_j)|^2 \rangle \\ \sigma_3 &= \max \left(\frac{\langle |\zeta_j F(\mathbf{x}; \mathbf{s}_j)|^3 \rangle}{\langle |\zeta_j F(\mathbf{x}; \mathbf{s}_j)|^2 \rangle} \right). \end{aligned}$$

In order to reproduce the exact covariance function, σ_2 is constrained to $\sigma_2^2 = 1$. Thus, one can maximize r under the constrain $\sigma_2 = 1$ and obtain through a simple functional

calculus that the ζ_j variables must take randomly the ± 1 values, i.e. their PDF is

$$p(\zeta) = \frac{1}{2}(\delta[\zeta + 1] + \delta[\zeta - 1]). \quad (6)$$

C. Canonically conjugated representations: Fourier & Blob case

The functions $F(\mathbf{x}; \mathbf{s})$ (2) are not unique, but are defined up to any unitary transformation $U(\mathbf{x}; \mathbf{s})$:

$$F'(\mathbf{x}; \mathbf{s}) = \int d\mathbf{s}_1 F(\mathbf{x}; \mathbf{s}_1) U(\mathbf{s}_1; \mathbf{s}) \quad (7)$$

$$\delta(\mathbf{s}_1 - \mathbf{s}_2) = \int d\mathbf{s} U(\mathbf{s}_1; \mathbf{s}) U(\mathbf{s}_2; \mathbf{s}), \quad (8)$$

where $F'(\mathbf{x}; \mathbf{s})$ is solution of Eq. (2). We note that $F(\mathbf{x}; \mathbf{s})$ can be interpreted either as an operator (the "square root" of the covariance operator $\mathcal{E}(\mathbf{x}; \mathbf{y})$) or as a set of parametric functions which reproduces the correlation.

A particularly important unitary transformation is the Fourier transform $U(\mathbf{s}_1; \mathbf{s}) = e^{i\mathbf{s}_1\mathbf{s}}$ which links two sets of **canonically conjugated representations**:

$$F'(\mathbf{x}; \mathbf{s}) = \int d\mathbf{s}_1 F(\mathbf{x}; \mathbf{s}_1) e^{i\mathbf{s}_1\mathbf{s}} \quad (9)$$

A particularly important representation is suggested by the relation (2) as scaled eigenvectors of the covariance operator $F(\mathbf{x}; \mathbf{s}) = \sqrt{\lambda(\mathbf{s})} \psi_{\mathbf{s}}(\mathbf{x})$ where $\psi_{\mathbf{s}}(\mathbf{x})$ is an eigenvector and $\lambda(\mathbf{s})$ its associated eigenvalue $\int d\mathbf{y} \mathcal{E}(\mathbf{x}; \mathbf{y}) \psi_{\mathbf{s}}(\mathbf{y}) = \lambda(\mathbf{s}) \psi_{\mathbf{s}}(\mathbf{x})$. This choice yields the Karhunen-Loeve representation [23].

We consider the case of homogeneous GRFs, $\mathcal{E}(\mathbf{x}; \mathbf{y}) \equiv \mathcal{E}(\mathbf{x} - \mathbf{y})$. The natural eigenvectors for a translation invariant operator are plane waves $\psi_{\mathbf{k}}(\mathbf{x}) = e^{i\mathbf{k}\mathbf{x}}$ while the corresponding eigenvalues $\lambda(\mathbf{k}) = S(\mathbf{k})$ where $S(\mathbf{k})$ is the spectrum, the Fourier transform of the covariance.

Thus, using the Karhunen-Loeve decomposition and searching for real-valued fields, we obtain a Fourier-like parametric function F_F . Choosing the transformation $U(\mathbf{a}; \mathbf{k}) = e^{i\mathbf{a}\mathbf{k}}$ one gets from F_F the "Blob-function" F_B :

$$\zeta F_F(\mathbf{x}; \mathbf{k}) \equiv \sqrt{S(\mathbf{k})} \sin(\mathbf{k}\mathbf{x} + \frac{\pi}{4}\zeta)$$

$$F_B(\mathbf{x}; \mathbf{a}) = \int d\mathbf{k} \sqrt{S(\mathbf{k})} e^{-i\mathbf{k}\mathbf{x}} e^{i\mathbf{k}\mathbf{a}} \equiv F_B(\mathbf{x} - \mathbf{a}).$$

Introducing these functions in the approximative discrete form derived (5) we obtain **the canonically conjugated Fourier and Blob representations**

$$\phi_F(\mathbf{x}) \approx L_k^{d/2} \sum_{j=1}^{N_c} \sqrt{S(\mathbf{k}_j)} \sin(\mathbf{k}_j \mathbf{x} + \frac{\pi}{4} \zeta_j) \quad (10)$$

$$\phi_B(\mathbf{x}) \approx L_a^{d/2} \sum_{j=1}^{N_c} \zeta_j F_B(\mathbf{x} - \mathbf{a}_j) \quad (11)$$

which differ from the discrete Fourier decomposition (FFT) or the discrete Moving-Average methods (MA) [24] through the use of stochastic wave numbers \mathbf{k}_j , blob positions \mathbf{a}_j and fixed phases $\zeta_j = \pm 1$.

The series (5) becomes finite if the functions $F(\mathbf{x}; \mathbf{s})$ have a compact support in the parametric space \mathbf{s} . Consequently, the number of terms in the sum N_c is roughly the ratio between the volume of the compact support and the chosen density of parameters ρ .

For the Fourier representation (10) the compact support is the domain in the reciprocal space $\{\mathbf{k}_i\}$ where the spectrum $S(\mathbf{k}_i)$ has non-negligible values. For the Blob representation (11) the compact support is the domain in the real space $\{\mathbf{a}_i\}$ where the blob function $F_B(\mathbf{x} - \mathbf{a}_i)$ has non-negligible values. Thus, these two methods require a similar number of terms N_c in the sum in order to calculate a realization of the field in a point \mathbf{x} with a given accuracy.

We note that the usual discrete Fourier decomposition (with fixed grid points) usually needs larger values of N_c , as demonstrated in the next section.

D. Advantages of the Fourier and Blob representations

The FFT approach is usually considered as one of the fastest construction techniques for GRFs. It allows one to compute the values of the field $\phi(\mathbf{x})$ on a physical, equidistant grid, of dimension N_g using N_g equidistant wavenumbers, with a numerical complexity $\mathcal{O}(N_g \log N_g)$. Using the Fourier/Blob methods (10),(11) with random $\{\mathbf{k}_i\}/\{\mathbf{a}_i\}$ to compute the values on the same grid requires a computational cost which scales as $\mathcal{O}(N_g \times N_c)$ where N_c is the number of parameters considered in the compact support. In general, we expect $\log N_g \ll N_c$. Even in this context, the proposed methods are particularly tempting because:

1. The randomness of the parameters improves the convergence rates toward Gaussianity, such that N_c is required to be only a few times larger than $\log N_g$ (especially in low dimensional spaces $d = 1, 2$).
2. The randomness of the parameters improves the way the parametric space is spanned, allowing for smooth convergent covariance for any number of terms in the series. FFT needs dense grids to achieve that.
3. The resulting fields are not spatially periodic.
4. The GRFs have a preserved structure of the equipotential lines (no interpolation needed for the field values in-between grid points).

The disadvantages are that the random parameters must be generated at every realization. The Blob method, requires a supplementary implementation of a nearest neighbor algorithm. The Blob method might not have always analytical Blob functions.

III. ACCURACY STUDY AND DNS TESTS

The Fourier (10) and Blob (11) representations are tested in the case of a 2D homogeneous GRF with the covariance

$$\mathcal{E}(x, y) \equiv \langle \phi(\mathbf{x}') \phi(\mathbf{x}' + \mathbf{x}) \rangle = \exp \left(-\frac{x^2}{2\lambda_x^2} - \frac{y^2}{2\lambda_y^2} \right), \quad (12)$$

which yields the associated Blob functions

$$F_B(x, y) = \sqrt{\frac{\pi\lambda_x\lambda_y}{2}} \exp \left(-\frac{x^2}{\lambda_x^2} - \frac{y^2}{\lambda_y^2} \right). \quad (13)$$

The correlation lengths are chosen $\lambda_x = 0.3$, $\lambda_y = 0.4$. The Fourier's compact support is a rectangle in which $|\mathbf{k}_i|\lambda_i \leq 5$ which ensures that 99.91% of the spectrum is reproduced. The Blob's compact support is a rectangle in which $|\mathbf{a}_i|/\lambda_i \leq 4$ which ensure that 99.99% of the covariance is reproduced.

More precisely, we investigate numerically the effects of the additional stochastic elements introduced in the representations (10) and (11) and the ability of the simple discrete distribution (6) to improve the convergence rate. For that, we shall use, further, six methods of computing the GRF, which are described in Table I. The notation for each type of

	FFC	FRC	FRD	BFC	BRC	BRD
Method	Fourier	Fourier	Fourier	Blob	Blob	Blob
Parameters	Fixed	Random	Random	Fixed	Random	Random
ζ	[0, 8)	[0, 8)	± 1	Gaussian	Gaussian	± 1

TABLE I: The numerical representations tested in this Section.

representation consists of three characters: the first letter for the method (Fourier or Blob), the second for how the parameters are distributed (Fixed or Random) and the third for the distribution of the random function ζ (Continuous or Discrete distributions).

A. Reproducing the covariance

We construct an ensemble of $M = 10^3$ realizations of the GRF $\phi(\mathbf{x})$ with the covariance (12) on a rectangular domain $[-\pi, \pi] \times [-\pi, \pi]$ using all six methods. A small number of parametric points $N_c = 12^2$ in the compact support was chosen for each method. The fluctuations of the resulting covariance around the exact profile $\delta\mathcal{E}(\mathbf{x}) = \langle \phi(\mathbf{0})\phi(\mathbf{x}) \rangle - \mathcal{E}(\mathbf{x})$ can be seen in Fig.1. All methods offer similar amplitudes except the FFC method, which, due to its fixed equidistant grid in the \mathbf{k} space has an unphysical periodicity.

The rate of convergence for the error of the covariance function $|\delta\mathcal{E}| = \int |\mathcal{E}(\mathbf{x}) - \mathcal{E}_{approx}(\mathbf{x})| d\mathbf{x}$ can be seen in Fig. 2 ($N_c = 12^2$) as function of M , the ensemble dimension. One can see that $|\delta\mathcal{E}|$ decays with the increase of M at approximately the same rate for five of the above methods and that the FFC (standard FFT) has a much weaker convergence at small values of N_c . These five methods are able to reproduce the covariance even at small values of the number of elements in the sums in Eqs. (10),(11). On the contrary, the FFC method (standard FFT) offers a poor representation of the covariance function on grids with low densities of points, in comparison with the other proposed methods. Increasing N_c , the decay rate of the error increases for the FFC method, but values similar to the other representation are attained at very large N_c (of the order $\sim 100^2$). Essentially, the fail is due to the weak stochastic character of FFC (fixed grid for the wave numbers). We note that the corresponding fixed grid Blob method (the BFC) gives much better results in spite of the same weak stochastic character.

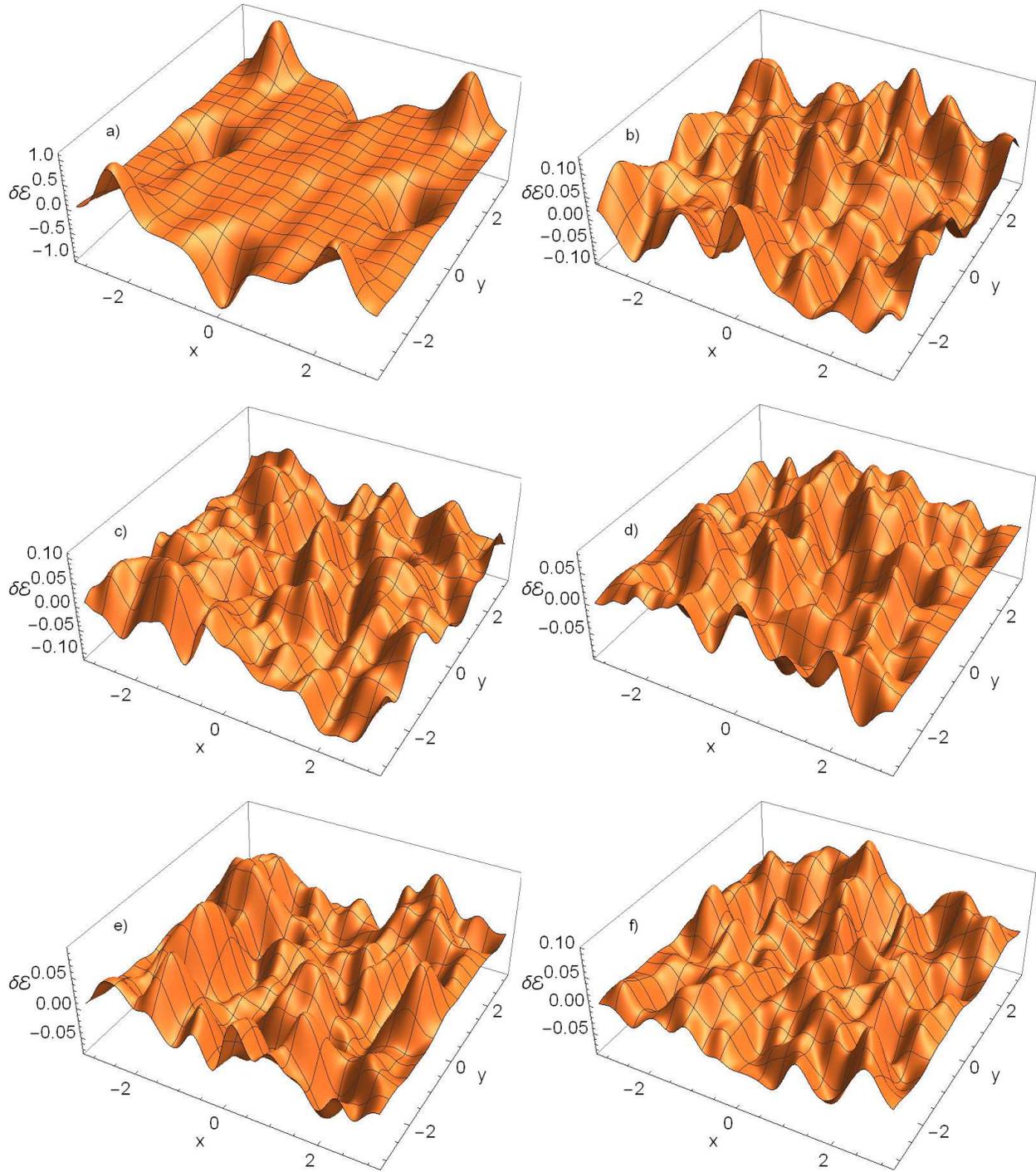


FIG. 1: The error $\delta\mathcal{E}(\mathbf{x})$ of the covariances averaged over ensembles of $M = 1000$ realizations with the FFC (a), FRD(b), FRC (c), BFC (d), BRD(e), BRC (f) methods.

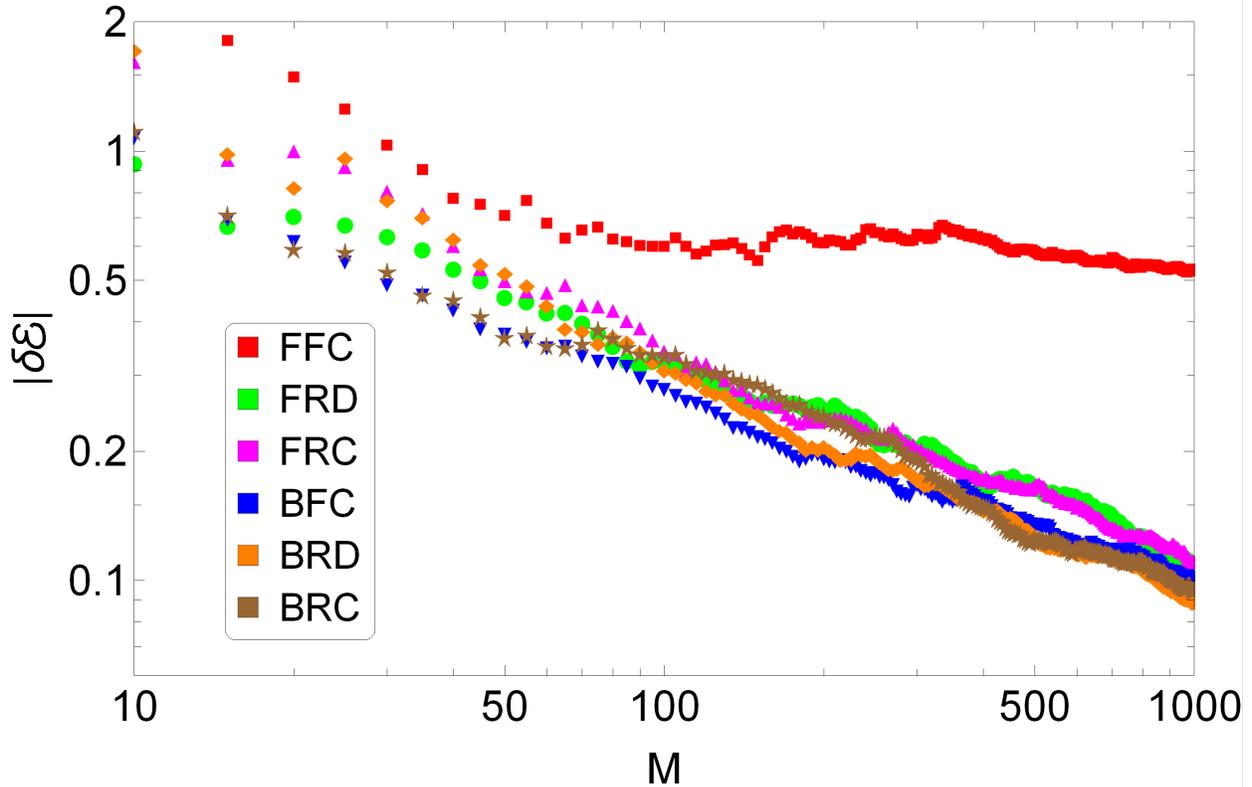


FIG. 2: Evolution of covariance error with the dimension of the ensemble for the FFC (red square), FRD (green circle), FRC (magenta up-triangle), BFC (blue down triangle), BRD (orange diamond), BRC (brown star) methods.

Thus, reasonable values of the error of the convolution are obtained with the FFC method at much larger values of M and/or N_c . The computational time that scales as $M \times N_c$ is much longer for the FFC than for the other five methods (by at least one order of magnitude).

B. Reproducing the Gaussian character

We have generated large ensembles ($M = 10^7$) of GRFs with all methods using even fewer points $N_c = 5^2$. In order to test the Gaussianity of the resulting fields, we have focused mainly on the one-point PDF of the field $\phi(\mathbf{0})$. We note that much larger values of M are necessary in order to reduce the statistical fluctuations in the computed PDFs. The results are presented in Fig. 3. One can see that the FFC has low quality for the potential distribution, as for the covariance function (Fig. 1). The corresponding Blob representation (with fixed grid, BFC) is even worse for $P(\phi)$ (see Fig. 3). It is obvious that the use of

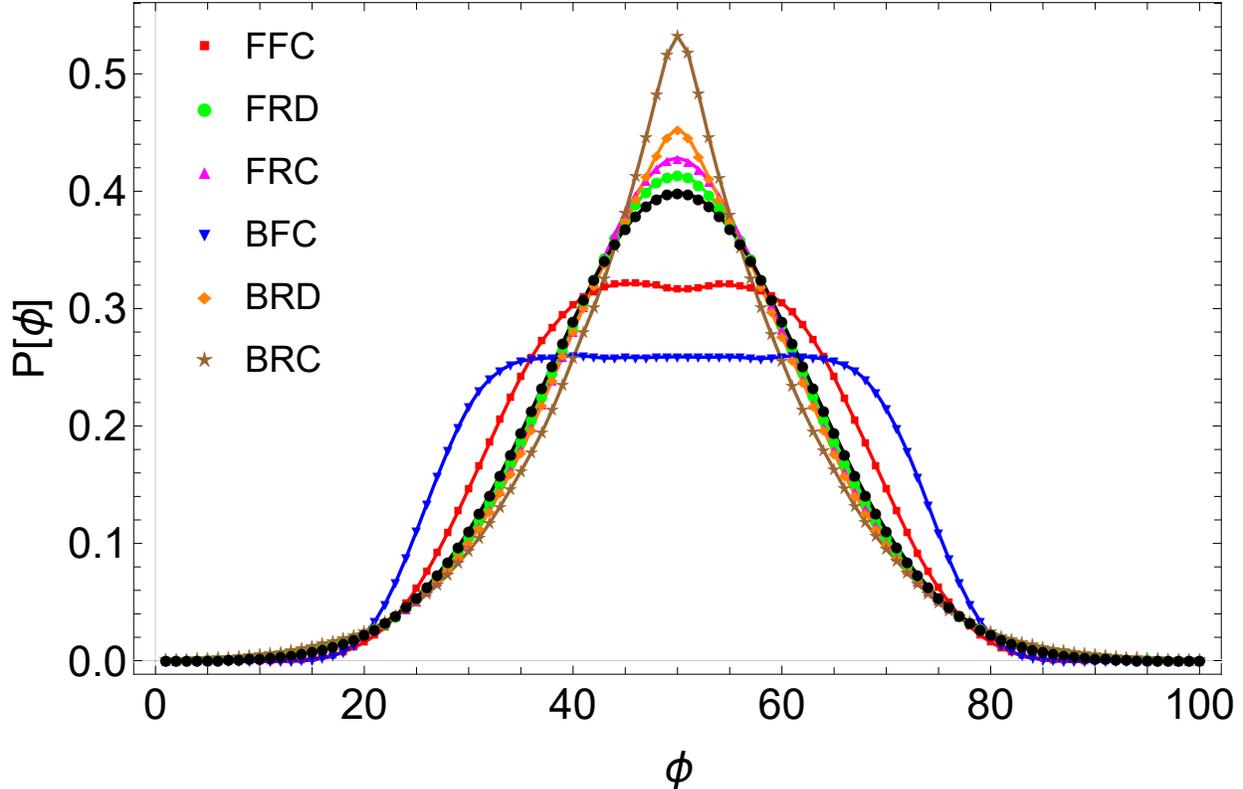


FIG. 3: PDFs of $\phi(\mathbf{0})$ obtained with the FFC (red square), FRD (green circle), FRC (magenta up-triangle), BFC (blue down triangle), BRD (orange diamond), BRC (brown star) methods and a statistical ensemble of $M = 10^7$ realizations.

random grids instead of fixed ones is a much better choice also in the matter of Gaussianity. Moreover, as it has been stated in Section II B, using discrete distributions $\zeta_j = \pm 1$, instead of distributions with continuous support, offers significant improvements in the profile $P(\phi)$: FRD and BRD are better than FRC and BRC.

Table II quantifies these results computing the first even moments for the PDF of $\phi(\mathbf{0})$. The global error defined as $|\delta P| = \int |P_{gauss}(\phi) - P_{method}(\phi)| d\phi$ and the 4-point correlation function $\mathcal{E}_{(4)} = \langle \phi(\mathbf{x}_1)\phi(\mathbf{x}_2)\phi(\mathbf{x}_3)\phi(\mathbf{x}_4) \rangle$ with $\mathbf{x}_1 = (0, 0)$, $\mathbf{x}_2 = (\lambda_x, \lambda_y)$, $\mathbf{x}_3 = (\lambda_x/2, -\lambda_y/3)$, $\mathbf{x}_4 = (-\lambda_x/3, \lambda_y/2)$ are also shown. One can see that fixed grids lead to sub-Gaussian distributions while random grids to over-Gaussian ones (longer tails).

Thus, we have shown that the best choices for the representation of homogeneous GRFs are based on random grids with $\zeta = \pm 1$, i.e. on FRD (10) or BRD (11) methods. Further, by Blob representation we shall refer to BRD while by Fourier to FRD methods in the remaining part of this paper. Note that the Fourier is slightly better than the Blob method.

	FFC	FRC	FRD	BFC	BRC	BR1	exact
$\langle \phi^2 \rangle$	1.046	1.002	0.999	1.264	1.001	1.001	1
$\langle \phi^4 \rangle$	2.504	3.424	3.223	3.257	3.976	3.543	3
$\langle \phi^6 \rangle$	8.447	21.32	18.38	11.04	30.83	23.37	15
$\langle \phi^8 \rangle$	35.13	199.82	145.48	44.19	372.89	235.67	105
$ \delta P $	0.167	0.061	0.026	0.309	0.136	0.066	0
$\mathcal{E}_{(4)}$	0.744	1.041	0.948	0.694	0.906	0.850	0.789

TABLE II: Numerical values of the first odd moments of the PDF $P(\varphi; \mathbf{0})$ for all six methods considered.

C. DNS of stochastic transport

We have proven until now that the Fourier (10) and Blob (11) representations offer the best convergence rates from the perspective of their Eulerian properties. Now, we perform additional tests regarding their Lagrangian abilities in the context of a DNS of a V-Langevin equation. The following model has been chosen:

$$\frac{d\mathbf{x}(t)}{dt} = \mathbf{v}(\mathbf{x}(t)) = \hat{e}_z \times \nabla\phi(\mathbf{x}(t)) + V_d\hat{e}_y, \quad (14)$$

where $\phi(\mathbf{x})$ is a GRF and $V_d\hat{e}_y$ is an average velocity. This stochastic equation describes the dynamics of test particles under electrostatic turbulence in magnetically confined plasmas [12, 25] or for tracer transport in incompressible turbulent fluids. The stochastic potential is considered frozen, i.e. the covariance is time independent. The covariance function is (12) with $\lambda_x = 1, \lambda_y = 2$.

We have chosen this transport model because the ensemble of solutions exhibits two invariants: a "local" one characteristic to each trajectory and a global one, characteristic to the entire ensemble. Both are a consequence of the null divergence $\nabla \cdot \mathbf{v}(\mathbf{x}) = 0$ property and of the homogeneity of the stochastic field. The equation of motion (14) is of Hamiltonian type, with $\phi_t(\mathbf{x}) = \phi(\mathbf{x}) + V_dx$ the Hamiltonian function. The latter is invariant in each realization of the potential $\phi(\mathbf{x})$, since the trajectories obtained from Eq. (14) evolve on the contour lines of $\phi_t(\mathbf{x})$. At $V_d = 0$, $\mathbf{x}(t)$ are closed and have periodic dependence on time, while at $V_d \neq 0$ some of the trajectories are opened.

The second invariant is statistical and involves the Lagrangian velocity $\mathbf{v}(\mathbf{x}(t))$. According to Lumley's Theorem [5, 26], the statistics of the Lagrangian velocity is identical with the statistics of the Eulerian velocity, at any time

$$P^L[\mathbf{v}(\mathbf{x}(t))] = P^E[\mathbf{v}(\mathbf{x})],$$

where $P^L = \langle \delta[\mathbf{v} - \mathbf{v}(\mathbf{x}(t))] \rangle$ is the Lagrangian probability and $P^E = \langle \delta[\mathbf{v} - \mathbf{v}(\mathbf{x})] \rangle$ is the Eulerian probability. The latter is a space-independent Gaussian function

$$P^E(\mathbf{v}) = \exp\left(-\frac{v_x^2}{2V_{xx}} - \frac{(v_y - V_d)^2}{2V_{yy}}\right),$$

where $V_{ii} = \langle v_i(\mathbf{0})v_i(\mathbf{0}) \rangle = -\partial_{jj}\mathcal{E}(\mathbf{x})|_{\mathbf{x}\rightarrow\mathbf{0}} = 1/\lambda_j^2$.

The existence of the constraints related to these invariants makes the transport process very complicated, but it also provides strong benchmarks for the numerical simulations.

Regarding the numerical implementation, a second order Runge-Kutta numerical integration scheme has been used for a time interval of $[0, t_{max}] = [0, 40]$ with a fixed time step $dt = 0.04$. An ensemble of $M = 3 \times 10^4$ trajectories has been resolved. We have implemented the Fourier (FRD) and Blob (BRD) representations with $N_c = 6^2, 12^2$. Two cases have been considered: $V_d = 0$ and $V_d = 0.4$. A Fourier simulation with N_c waves is denoted by FN_c while a Blob one by BN_c where $N_c = 36$ or 144 .

We underline that M , the dimension of the ensemble and (especially) N_c , the number of random parameters in the series (10), (11), are small compared to the usual values in DNS. Thus, the DNS can be performed on personal computers, where the typical running times are rather small, of the order of $t_{CPU} \sim 10^2 s$.

First, we have checked that the numerical integration and the use of the generators (10),(11) of the GRF do not affect the Hamiltonian character of the trajectories. We plot in Fig. 4 a randomly chosen trajectory for almost 10 times its period. Qualitatively, the trajectory remains closed. Apart from small oscillations $\delta\phi(t)/\bar{\phi} \sim 10^{-4}$, the potential is perfectly conserved along the represented trajectory. Thus, the combined errors from the approximation of the field and from the numerical integration remain small.

Second, we test the global invariants by computing the PDF of Lagrangian velocities $P^L[\mathbf{v}(\mathbf{x}(t))]$ as well as its first moments $\langle v_i^j(t) \rangle$, $j = 1, 4$. Figure 5 shows the components of this distribution of the Lagrangian velocity at the moment $t = 20$, compared with the exact, Gaussian profiles. The results are close to the theoretical distributions, even for the

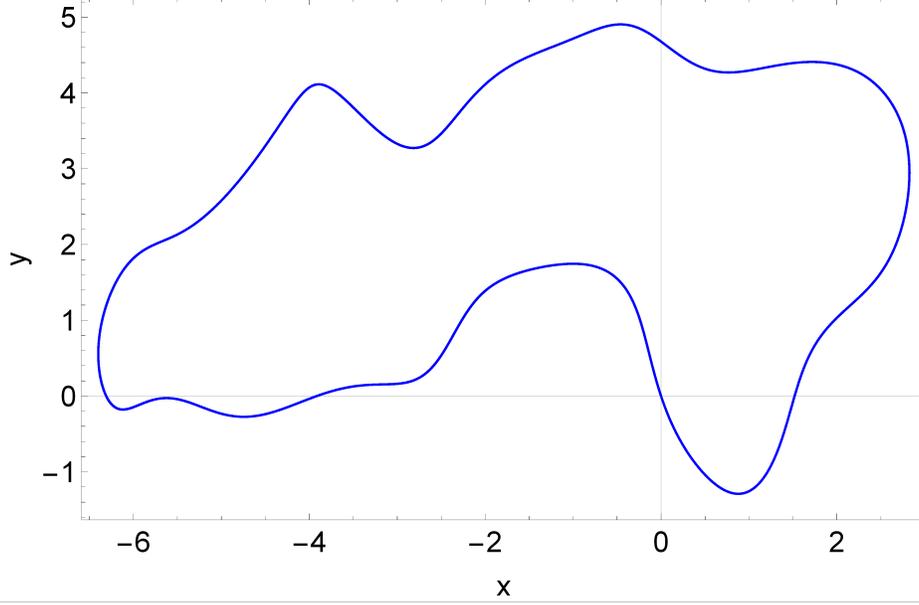


FIG. 4: A typical trajectory simulated over 10 periods.

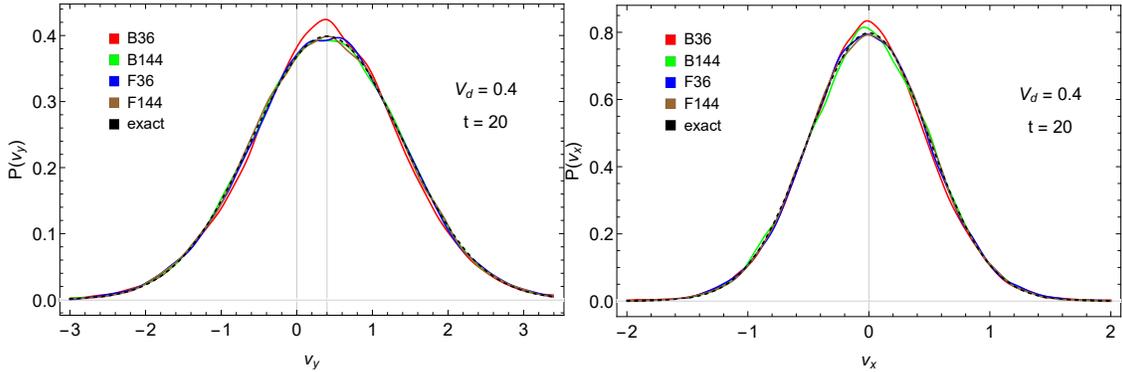


FIG. 5: Large time $t = 20$ distribution of velocities. The exact, Gaussian shapes are in dashed lines.

cases $N_c = 36$. The statistical fluctuations can be analyzed more clearly in Figure 6 where the moments $\langle v_i^j(t) \rangle$ are shown. On average, the Lagrangian invariance is well reproduced by both methods, the fluctuations being a consequence of a finite ensemble ($M = 3 \times 10^4$). The deviations of the average values (for example $\langle \bar{v}_y^4 \rangle \approx 3.3$ instead of the exact value 3 for $B36$ method) are a consequence of a finite N_c . As seen in Figure 6, the increase of N_c approaches the averages to the theoretical values and reduces the statistical fluctuations. The results are satisfactory even at the small values taken here.

A property of the stochastic transport described by Eq. (14) is that the Gaussian,

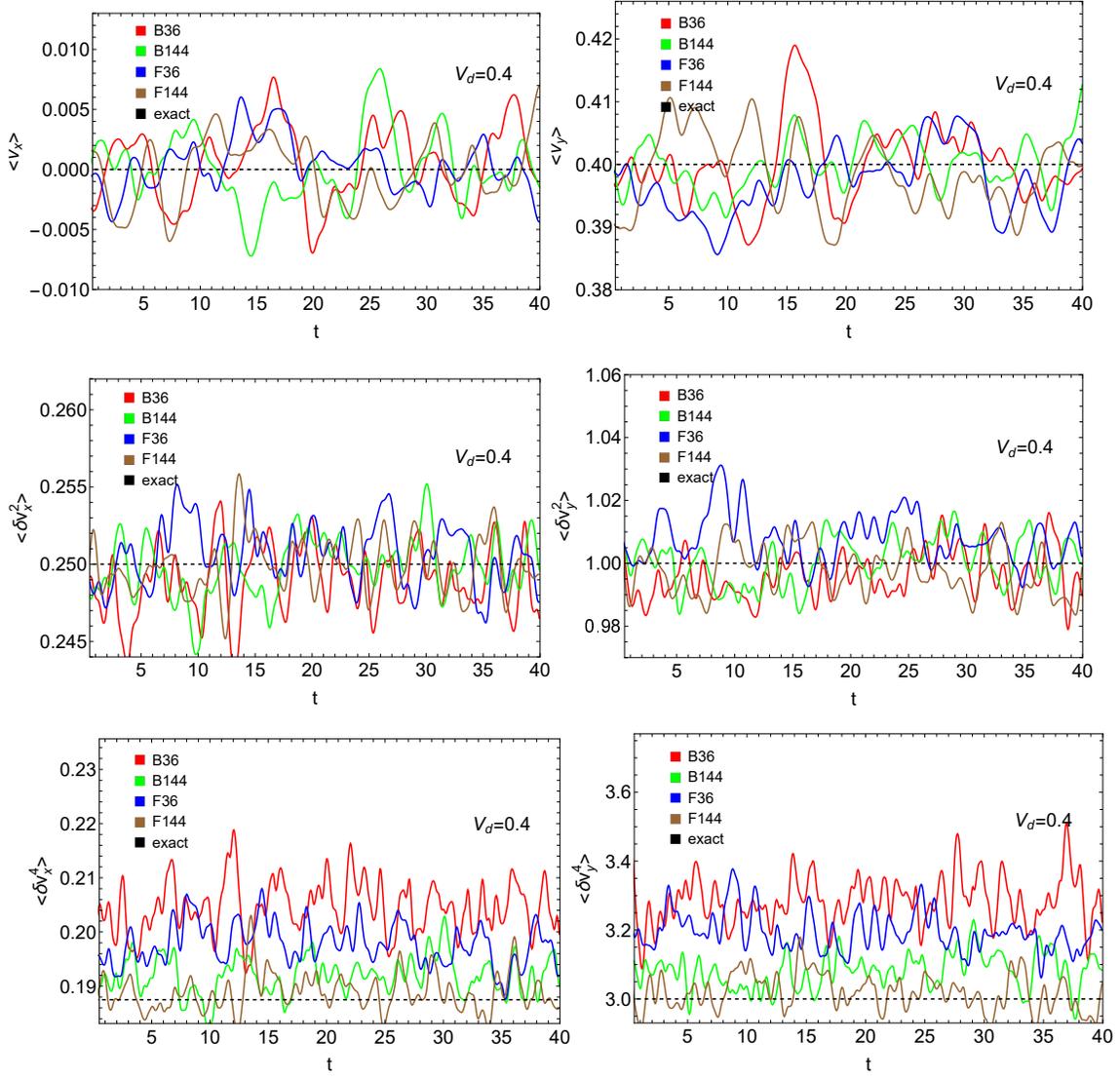


FIG. 6: Time dependence of average Lagrangian velocity moments.

time-invariant Lagrangian velocity does not yield a Gaussian distribution of the trajectories $P(\mathbf{x}(t)) = \langle \delta[\mathbf{x} - \mathbf{x}(t)] \rangle$. The latter is a peaked function with long tails [27]. This happens because of the trajectory trapping or eddying, produced by the invariance of the Lagrangian potential that ties particle paths on its contour lines [25, 28]. An average velocity opens a part of trajectories along its direction (\hat{e}_y), but trapped particles still exist [29], as seen in Fig. 7 for $V_d = 0.4$.

There are no clear theoretical results on $P(\mathbf{x}(t))$ that could be used as benchmark of the present DNS. Instead, we compare the results of the four runs commented here. The probability of the displacements $x(t)$ and $y(t)$ is shown in Fig 7. The only observable

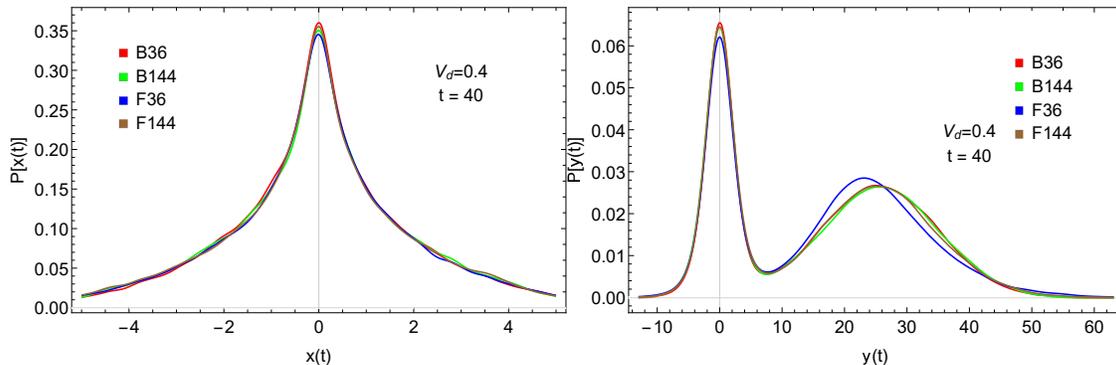


FIG. 7: Projected trajectory distributions $P(x(t)), P(y(t))$

difference appears in the distribution $P[y(t)]$ obtained in the $F36$ run. It underestimates the average displacement, the spreading of the free trajectories as well as the number of trapped particles.

The time dependent diffusion coefficients $2D_{ii}(t) = d\langle x_i^2(t) \rangle / dt$ are presented in Fig. 8. The decay of D_{ii} at large time is the consequence of trajectory trapping. One can see that all simulations yield practically the same result at $t < 1$ and that significant differences appear at $t > 1$, especially between the calculations at $N_c = 36$ and those at $N_c = 144$. This figure also shows that the $F36$ method underestimates the trapping of particles (it yields a slower decay of D_{ii} at large times $t > 2 - 3$). The converse is true for the $B36$ method which overestimates the trapping by smaller values of D_{ii} at large times. All methods give a dependence of the diffusion coefficients as $\mathcal{D} \propto t^\gamma$. For $F36$ $\gamma \approx -0.15$ while for $B36$ $\gamma \approx -0.4$. At larger values of N_c , $F144$ and $B144$, we can observe how the results and the slope γ converge towards a common profile with $\gamma \approx -0.3$, in accordance with well known results [30].

D. The hybrid representations

The results from Fig. 8 with no bias $V_d = 0$ suggest that there are some intrinsic pathologies within the Fourier and Blob representations. The first seems to produce very long (quasi-free) trajectories, while the latter very small, closed and less complex trajectories. The explanation is related to the specific form of the parametric functions of each representation. Using waves (Fourier) with a small number of terms is more likely to pro-

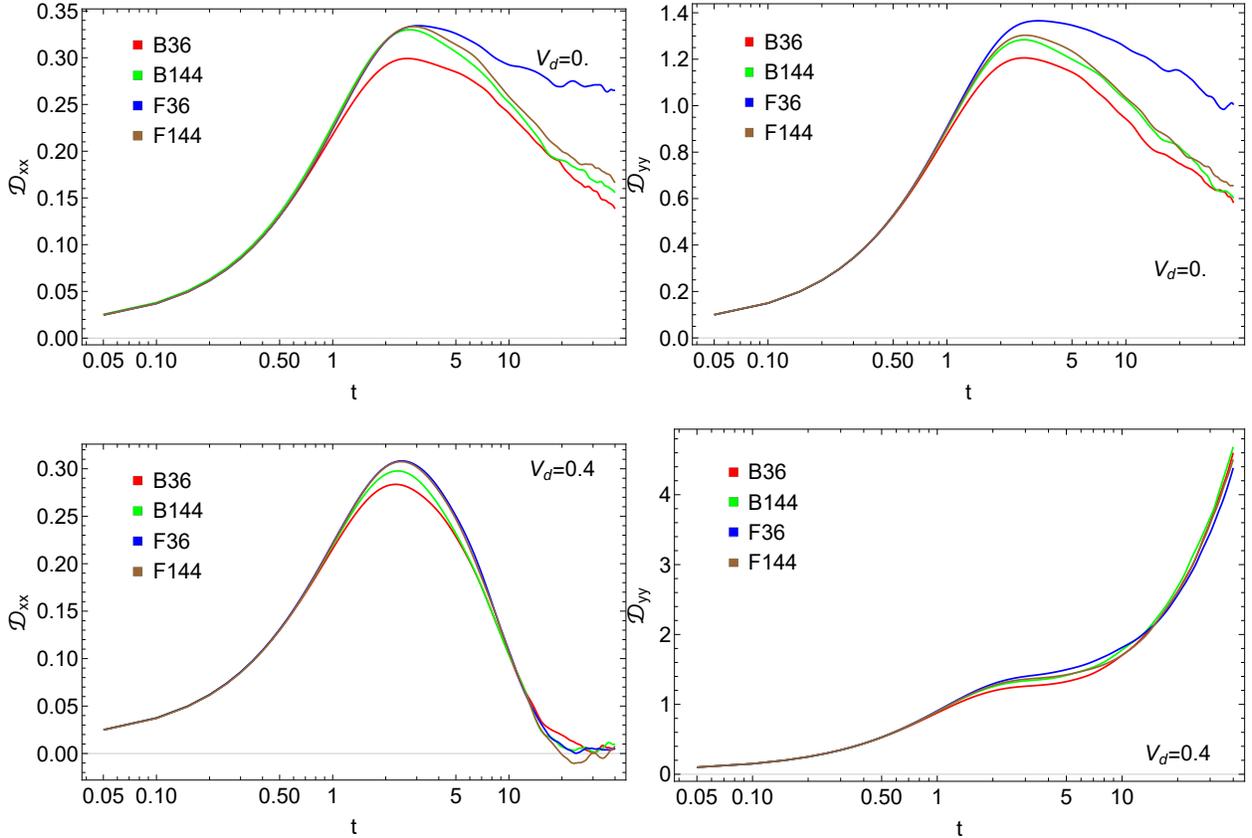


FIG. 8: Diffusion coefficients $\mathcal{D}_{ii}(t)$ obtained with the methods B36 (red), F36 (blue), B144 (green), F144 (brown).

duce long equipotential lines (which are, in fact, trajectories). A single plane wave is unable to produce a closed field line. In contrast, even a single Blob function will generate an inherently closed trajectory. A small number of Blob functions is unlikely to produce long equipotential lines.

Also, the Table II suggests that the Blob method reproduces better than the Fourier method the higher-order correlations ($\mathcal{E}_{(4)}$). The overestimation of these correlations corresponds to smoother fields, which means less complex fields. But a less complex field has less complex equipotential lines and, consequently, less complex Lagrangian solutions. The overestimation of the correlation in the Fourier representation is natural: the waves are omnipresent, thus, any two points are "correlated" through the wave. Only the statistical averaging of the phases can decouple them.

These shortcomings do not affect the Lagrangian distribution of velocities or the average velocity of the ensemble, as seen in Figs. 5,6. Their effect is visible in the diffusion

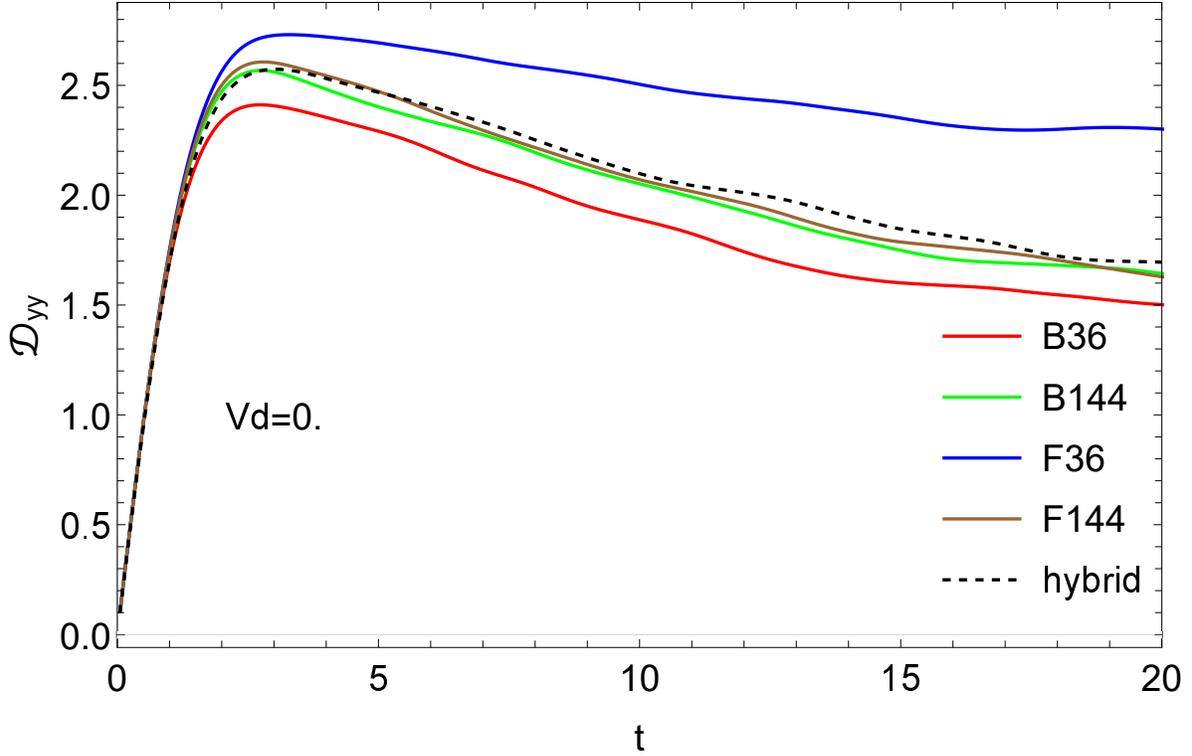


FIG. 9: The diffusion coefficient D_{yy} obtained with the Fourier-Blob representation (15) with $N_c = 36$ (for both Fourier and Blob terms) compared to the results presented in Fig. 8

coefficients, which show a much stronger dependence on N_c .

These structural properties of the Blob and Fourier methods can be exploited to yield improved results for the diffusion coefficients without increasing N_c . We propose a hybrid representation of the GRF that combines the Fourier and Blob methods:

$$\phi_{FB}(\mathbf{x}) = \eta_1 \phi_F(\mathbf{x}) + \eta_2 \phi_B(\mathbf{x}). \quad (15)$$

with $\eta_1^2 + \eta_2^2 = 1$. We show that the systematic errors of the two methods compensate in this Fourier-Blob (FB) representation. The results obtained for $\eta_1 = \eta_2 = 1/\sqrt{2}$ using the Fourier-Blob approach (15) with $N_c = 36$ are shown in Fig. 9 for the diffusion coefficient $D_{yy}(t)$. The resulting profile is very close to the profiles obtained with $F144$ and $B144$ simulations at any time.

Finally, it must be emphasized that the hybrid approach enables the possibility of reproducing accurate Lagrangian properties of stochastic transport while requiring roughly half of the CPU time required by the use of only Fourier or Blob representations. A factor of

two becomes highly relevant when dealing with more complex covariance functions which may require *a priori* a larger number of parametric functions N_c .

IV. SUMMARY AND CONCLUSIONS

The general integral representation of the GRFs (1),(2), contains a parametric function $F(\mathbf{x}; \mathbf{s})$ and an uncorrelated random variable $\zeta(\mathbf{s})$. We have derived from Eqs. (1),(2) a set of discrete representations. They are of Blob and Fourier type, according to the parametric function that is a space structure $F_B(\mathbf{x} - \mathbf{a}_j)$ in the first case and a wave amplitude structure in the second case $F_F(\mathbf{x}; \mathbf{k}_j) = \sqrt{S(\mathbf{k}_j)} \sin(\mathbf{k}_j \mathbf{x})$. Additional stochastic elements were introduced in both types of representations by considering the points \mathbf{a}_j and the wave numbers \mathbf{k}_j as stochastic parameters with uniform distributions. The random variable ζ was taken with discrete ($\zeta = \pm 1$) support.

Six representations of the GRF, defined in Table I, were analyzed to prove that our proposal Fourier (FRD) and Blob (BRD) me provide a better convergence of the Eulerian properties than other standard representations. We have shown that reasonable errors in the covariance and in the PDF of the potential are obtained at much smaller values of N_c and M than in the usual Fourier representation (FFC). This leads to the decrease of the computing times by at least one order of magnitude compared to the usual FFC method.

The convergence of the Lagrangian properties of these two methods were further analyzed in the frame of the DNS of a special type of stochastic transport described by a V-Langevin equation in two-dimensional, time-independent velocity fields with zero divergence. The invariance of the Lagrangian potential in each realization and the statistical invariance of the Lagrangian velocity provide benchmarks for the validation of the numerical results. We have shown that simulations with both Fourier and Blob methods satisfy these constraints with good precision for $N_c \gtrsim 100$ and $M \gtrsim 10^4$. The main difference between these representations appear in their ability to describe the effects of trajectory trapping or eddying on the contour lines of the potential.

The Fourier (FRD) results underestimate while the Blob (BRD) method overestimates the effects of trapping on the diffusion coefficients. These systematic errors were strongly reduced by a hybrid representation which combines linearly the Fourier and Blob series in a single Fourier-Blob method. The result is a representation able to decrease the value of

N_c required for a certain accuracy and such to reduce the calculation time by a factor 2 compared to the BRD and FRD.

In conclusions, we have strongly improved the representation of the GRFs by introducing additional random elements. We have shown that the hybrid Fourier-Blob method (15) provides a fast tool that can be used in the numerical studies of complex stochastic advection processes. This opens the possibility of performing such studied on personal computers. For the case analyzed here, typical running times are of the order of 10^2s or even less for the hybrid representation.

V. ACKNOWLEDGEMENT

This work has been carried out within the framework of the EUROfusion Consortium and has received funding from the Euratom research and training programme 2014-2018 under grant agreement No 633053 and also from the Romanian Ministry of Research and Innovation. The views and opinions expressed herein do not necessarily reflect those of the European Commission.

-
- [1] Kampen. *Stochastic processes in physics and chemistry*. Elsevier, Amsterdam Boston London, 2007.
 - [2] Paul Bressloff. *Stochastic processes in cell biology*. Springer, Cham, 2014.
 - [3] Wolfgang Paul. *Stochastic processes : from physics to finance*. Springer, Berlin New York, 2013.
 - [4] Andreas Diekmann and Peter Mitter. *Stochastic modelling of social processes*. Academic Press, 2014.
 - [5] A. S. Monin. *Statistical fluid mechanics : mechanics of turbulence*. MIT Press, Cambridge, Mass, 1971.
 - [6] R. Balescu. V-langevin equations, continuous time random walks and fractional diffusion. *Chaos, Solitons & Fractals*, 34(1):62 – 80, 2007. In Search of a Theory of Complexity.
 - [7] Radu Balescu. *Aspects of anomalous transport in plasmas*. CRC Press, Place of publication not identified, 2005.

- [8] B. Ganapathysubramanian and N. Zabaras. A stochastic multiscale framework for modeling flow through random heterogeneous porous media. *Journal of Computational Physics*, 228(2):591 – 618, 2009.
- [9] Yang Liu, Jingfa Li, Shuyu Sun, and Bo Yu. Advances in gaussian random field generation: a review. *Computational Geosciences*, 23(5):1011–1047, Oct 2019.
- [10] MARC BOIVIN, OLIVIER SIMONIN, and KYLE D. SQUIRES. Direct numerical simulation of turbulence modulation by particles in isotropic turbulence. *Journal of Fluid Mechanics*, 375:235263, 1998.
- [11] G. Manfredi and R. O. Dendy. Test-particle transport in strong electrostatic drift turbulence with finite larmor radius effects. *Phys. Rev. Lett.*, 76:4360–4363, Jun 1996.
- [12] J.-D. Reuss and J. H. Misguich. Low-frequency percolation scaling for particle diffusion in electrostatic turbulence. *Phys. Rev. E*, 54:1857–1869, Aug 1996.
- [13] Tijana Radivojević and Elena Akhmatskaya. Modified hamiltonian monte carlo for bayesian inference. *Statistics and Computing*, 30(2):377–404, 2020.
- [14] V. Naulin, A. H. Nielsen, and J. Juul Rasmussen. Dispersion of ideal particles in a two-dimensional model of electrostatic turbulence. *Physics of Plasmas*, 6(12):4575–4585, 1999.
- [15] Di Yang and Lian Shen. Direct numerical simulation of scalar transport in turbulent flows over progressive surface waves. *Journal of Fluid Mechanics*, 819:58–103, 2017.
- [16] Petter Abrahamsen. *A review of Gaussian random fields and correlation functions*. Norsk Regnesentral/Norwegian Computing Center, Oslo, 1997.
- [17] Francisco Cuevas, Denis Allard, and Emilio Porcu. Fast and exact simulation of gaussian random fields defined on the sphere cross time. *Statistics and Computing*, 30(1):187–194, 2020.
- [18] Arno Solin and Simo Särkkä. Hilbert space methods for reduced-rank gaussian process regression. *Statistics and Computing*, 30(2):419–446, 2020.
- [19] R. C. Tautz and A. Dosch. On numerical turbulence generation for test-particle simulations. *Physics of Plasmas*, 20(2):022302, 2013.
- [20] R.C. Tautz. On simplified numerical turbulence models in test-particle simulations. *Journal of Computational Physics*, 231(14):4537 – 4541, 2012.
- [21] T. Hauff and F. Jenko. Turbulent eb advection of charged test particles with large gyroradii. *Physics of Plasmas*, 13(10):102309, 2006.

- [22] V. Yu. Korolev and I. G. Shevtsova. On the upper bound for the absolute constant in the berryseen inequality. *Theory of Probability & Its Applications*, 54(4):638–658, 2010.
- [23] Yingbo Hua and Wanquan Liu. Generalized karhunen-loeve transform. *IEEE Signal Processing Letters*, 5(6):141–142, June 1998.
- [24] Mickaële Le Ravalec, Benoît Noetinger, and Lin Y. Hu. The fft moving average (fft-ma) generator: An efficient numerical method for generating and conditioning gaussian simulations. *Mathematical Geology*, 32(6):701–723, Aug 2000.
- [25] M. Vlad, F. Spineanu, J. H. Misguich, and R. Balescu. Diffusion in biased turbulence. *Phys. Rev. E*, 63:066304, May 2001.
- [26] James P. Gleeson. Comment on “diffusion in biased turbulence”. *Phys. Rev. E*, 66:038301, Sep 2002.
- [27] Madalina Vlad and Florin Spineanu. Trajectory structures and transport. *Phys. Rev. E*, 70:056304, Nov 2004.
- [28] M. Vlad, F. Spineanu, J. H. Misguich, and R. Balescu. Diffusion with intrinsic trapping in two-dimensional incompressible stochastic velocity fields. *Phys. Rev. E*, 58:7359–7368, Dec 1998.
- [29] M Vlad and F Spineanu. Random and quasi-coherent aspects in particle motion and their effects on transport and turbulence evolution. *New Journal of Physics*, 19(2):025014, feb 2017.
- [30] M. B. Isichenko. Percolation, statistical topography, and transport in random media. *Rev. Mod. Phys.*, 64:961–1043, Oct 1992.

