

Rare-event sampling applied to the simulation of extreme mechanical efforts exerted by a turbulent flow on a bluff body

Thibault Lestang^{1,2†}, Freddy Bouchet¹ and Emmanuel Lévéque²

¹Univ Lyon, ENS de Lyon, Univ Claude Bernard de Lyon, CNRS, Laboratoire de Physique, F-69342 Lyon, France

²Univ Lyon, Ecole Centrale de Lyon, Univ Claude Bernard de Lyon, INSA de Lyon, CNRS, Laboratoire de Mécanique des Fluides et d'Acoustique, F-69134 Ecully cedex, France

(Received xx; revised xx; accepted xx)

This study evaluates the relevance of rare-event sampling techniques to accelerate the simulation of extreme mechanical efforts exerted by a turbulent flow impinging onto a bluff body. The main idea is to replace a long simulation by a set of much shorter ones, running in parallel, with dynamics that are replicated or pruned in order to sample large-amplitude events more frequently. Such techniques have been shown to be efficient for a wide range of problems in statistical physics, computer science, biochemistry, enabling the simulation of rare events otherwise out of reach by direct sampling. This work is the first application to fluid-structure interaction problems. The drag experienced by a squared obstacle placed in a turbulent flow (in two dimensions) is taken as a representative case study to investigate the performance of two major rare-event sampling algorithms, namely the Adaptive Multilevel Splitting (AMS) and the Giardina-Kurchan-Tailleur-Lecomte (GKTL) algorithms. Practical evidence is given that the fast sweeping-time of fluid structures past the obstacle has a drastic influence on the efficiency of these two algorithms. While it is shown that the AMS algorithm does not yield significant run-time savings, the GKTL algorithm appears to be efficient to sample extreme fluctuations of the time-averaged drag and estimate related statistics such as return times. Beyond the study of applicability of rare-event sampling techniques to a fluid-mechanical problem, this work also includes a detailed phenomenological description of extreme-drag events of a turbulent flow on a bluff body.

1. Introduction

Turbulent flows are important in a variety of natural phenomena, industrial and civil applications. Their characteristic feature is the spontaneous development of intense and sporadic motions associated with extreme internal forces (Lesieur 2011; Donzis & Sreenivasan 2010; Yeung *et al.* 2015). ‘‘Extreme’’ refers here to fluctuations that can deviate from the mean value by $\mathcal{O}(10)$ standard deviations. In engineering, the nature of such extreme dynamical events and their statistics are of crucial interest to predict excessive mechanical efforts. Such anomalous constraints can threaten the structural integrity of embedded structures.

From the viewpoint of chaotic dynamical systems, turbulence in fluids is linked to non-linearity and strong departure from statistical equilibrium (Kraichnan & Chen 1989). The use of perturbative methods in identifying resonant interactions among degrees of freedom responsible for extreme fluctuations is unsuccessful. Therefore, simulation offers

† Email address for correspondence: thibault.lestang@cs.ox.ac.uk

a practical approach to gain physical insight into these events, quantifying their intensity and estimating their frequency of occurrence. However, this requires very long simulations since these events are rare. Rare-event sampling refer to a large body of methods that aim at exploring preferentially regions of phase space corresponding to rare events, that would otherwise be accessed with a very low probability through a brute-force direct sampling. In the present work, a computational study of extreme mechanical efforts acting on an immersed bluff body is conducting by using both very long time-series (direct sampling) and rare-event sampling techniques.

In fluid turbulence, rare-event sampling has been approached mainly from the perspective of simplified dynamics such as the one-dimensional Burgers' equation with a stochastic forcing (Bec & Khanin 2007). In this case, dynamics can be sampled by using a Markov chain Monte-Carlo algorithm (Düben *et al.* 2008; Mesterházy & Jansen 2011; Mesterházy *et al.* 2013) that provides a framework for rare-event sampling. An alternative approach is based on instantons (Gurarie & Migdal 1996; Grafke *et al.* 2015) and applies to stochastically driven systems in the limit of weak noise. Instantons refer to the most probable trajectories in phase space that achieve a given rare event (in the limit of weak noise). Suitable numerical schemes can be used to evaluate instantons as well as the related probabilities of rare events (Chernykh & Stepanov 2001; Grafke *et al.* 2013; Grigorio *et al.* 2017; Laurie & Bouchet 2015; Bouchet *et al.* 2014). An example is the investigation of the physics of rogue waves (Dematteis *et al.* 2018, 2019). A drawback of the aforementioned approaches is their limitation to simple and stochastically driven dynamics.

In this paper, a more general approach is considered for complex, possibly deterministic, dynamical systems. It is based on sampling algorithms relying on *selection rules* applied to an ensemble of trajectories, and is designed to sample rare events of some observable with a higher frequency. Even though such ideas date back to the early 1950s, they have received ever-growing interest over the the last twenty years with successful applications in various domains such as chemistry (van Erp & Bolhuis 2005; Escobedo *et al.* 2009; Teo *et al.* 2016), biophysics (Huber & Kim 1996; Zuckerman & Chong 2017; Bolhuis 2005), nuclear physics (Louvin *et al.* 2017), nonlinear dynamical systems (Tailleur & Kurchan 2007) and communication networks simulation (Villen-Altamirano & Villen-Altamirano 1994). More importantly, these type of algorithms have been shown to be useful for the study of rare events in simple deterministic dynamics (Wouters & Bouchet 2016). An original contribution of the present work is certainly the application of rare-event sampling algorithms in the context of far-from-equilibrium dynamics with an irreducible very large number of degrees of freedom. Two different algorithms suitable for out-of-equilibrium dynamics are considered and compared. Namely, the Adaptive Multilevel Splitting algorithm and the Giardina-Kurchan-Tailleur-Lecomte algorithm.

The Adaptive Multilevel Splitting algorithm (Cérou & Guyader 2007) builds on previous ideas about splitting approaches (Kahn & Harris 1951; Glasserman *et al.* 1998, 1999) – a detailed description will be given later. In recent years, it has allowed for the computation of rare events in problems involving a large number of degrees of freedom such as molecular dynamics simulations (Aristoff *et al.* 2015; Teo *et al.* 2016). The first use of the AMS algorithms for more complex dynamics, for instance stochastic partial differential equations was for the computation of rare trajectories in the Allen-Cahn equations (Rolland *et al.* 2016). More recently is has been applied to rare events in stochastic models of wall-turbulence (Rolland 2018) and atmospheric dynamics (Bouchet *et al.* 2019).

During the last decade, the main theoretical framework for the study of rare events in statistical physics has been the theory of large deviations (Touchette 2009). Alongside

numerical methods have been developed to sample rare events (Moral 2004a). Among them the GKTL algorithm (Giardinà *et al.* 2006) is particularly suited for (chaotic) dynamical systems (Giardinà *et al.* 2011; Laffargue *et al.* 2013). Recently the GKTL algorithm has allowed to successfully perform the numerical simulation of extreme heat waves in a simplified modelling of the atmosphere (Ragone *et al.* 2018). This achievement represents a significant leap in the applicability of rare-event sampling to complex dynamical systems. Along the same line, rare-event sampling algorithms are here pushed aside traditional applications to consider fluid-structure interaction in a turbulent flow.

The paper is organized in two parts. The first part highlights the phenomenology of extreme fluctuations of the drag force acting on a square placed in a two-dimensional turbulent channel flow. This study is based on the simulation of the flow over a very long duration, made possible by the relative simplicity of the flow. Motivation for this study is twofold. Firstly, it provides a detailed description of the statistics and dynamics related to extreme drag fluctuations. This analysis is informative from the viewpoint of fluid mechanics and, to the best of our knowledge, has never been reported before. Secondly, it yields reference results that are required to validate the outputs of rare-event algorithms and to evaluate the possible gain obtained from them. This assessment is developed in the second part of the paper.

In section 2, the flow set-up is introduced and the dynamics related to typical drag fluctuations is described. The statistical properties of the drag are then discussed. In section 3, the phenomenology of extreme fluctuations of extreme drag fluctuations is investigated based on a direct sampling approach. Both the instantaneous drag and time-averaged drag are considered. It is found that sampled extreme events for the instantaneous drag share very similar dynamics. Furthermore, extreme fluctuations for the time-averaged drag can be connected to the statistics of the instantaneous drag. This feature is well supported by theoretical arguments applied to simplified stochastic dynamics. Section 4 reports the applicability of both the AMS and GKTL algorithms to the numerical simulation of extreme drag fluctuations, by using the same flow configuration. In section 4.1, we show that the use of the AMS algorithm is not successful, or at least not straightforward. This difficulty is put in perspective with the phenomenology developed in the previous sections. Section 4.2 presents the computation of extremes of the time-averaged drag, using the GKTL algorithm. This latter allows an exceptional reduction of the computational cost required to simulate trajectories corresponding to extreme time-averaged drag values. As a specific application, the GKTL algorithm is used to compute the return times of extreme fluctuations of the time-averaged drag acting on the immersed obstacle. Perspectives and conclusion end this work.

2. Description of the numerical case study

The drag exerted by a grid-generated turbulent flow onto a fixed squared obstacle is considered as a representative case study (see Fig. 1). Although real-world applications would eventually imply three-dimensional dynamics, a simplified two-dimensional setting has been chosen here to reduce the computational cost and allow for a systematic study. We believe that this system embeds the characteristic features that makes the application of rare-event algorithms both relevant and challenging for fluid-structure-interaction problems. Turbulent eddies generated in the near-wake of the grid are carried downstream. They interact with each other and grow in size as expected for two-dimensional turbulent dynamics. The dimension of the grid is such that the size of the eddies that hit the square is comparable to its size, resulting in strong fluctuations of the drag acting on the square.

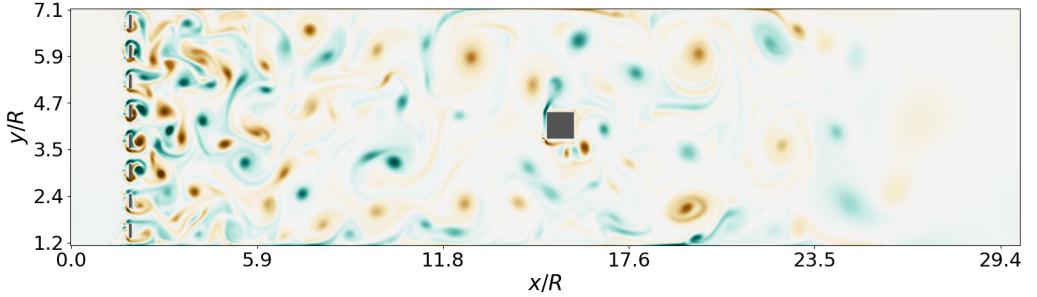


Figure 1: Our case study is a grid-generated turbulent flow impinging onto a fixed squared obstacle (of size R) located at the centre of a channel in two dimensions. The flow is artificially damped near the end of the channel. In the developed flow, turbulent eddies have typically the size of the square, which results in strong fluctuations of mechanical efforts acting on the square. The vorticity is displayed with an arbitrary colour map from blue (negative values) to red (positive values).

The flow dynamics is integrated by the Lattice Boltzmann Method (LBM) in our numerical simulations. While traditional methods in computational fluid dynamics rely on a discretization of the Navier-Stokes equations, the LBM considers the fluid at a mesoscopic level. Capturing the dynamics of collections of fluid particles distributed on a lattice is here preferred to solving non-linear PDEs. Further details about the LBM are given in Appendix and references therein. In our context, this numerical method has been chosen principally for its computational manageability and efficiency.

The simulated flow develops in a long plane channel of dimension 513×129 mesh points. The square obstacle has size $R = 16$ (in mesh unit) and is located at the centre of the channel. The spacing and bar height of the entrance grid are both equal to $R/2$ (see Fig. 1). No-slip boundary conditions are enforced on top and bottom walls of the channel and on the surface of the obstacle by using an halfway bounce-back procedure (Sukop & Jr 2006). Upstream of the grid, a constant parabolic velocity profile and a constant mass density (equal to unity) are imposed as an inlet condition. The centerline velocity is 0.05 in lattice units, *i.e.* normalised by Δx and Δt referring to the lattice resolution and the time-step respectively. The initial distributions are imposed at equilibrium (see Appendix). In the bulk, the viscosity is adjusted so that grid turbulence is generated with Reynolds number $\text{Re}_{\text{grid}} = 1200$. The reference Mach number is equal to 0.06 in agreement with the assumption of weak compressibility of the LBM. Near the end of the channel, the flow is progressively damped within a *sponge layer* where the viscosity is artificially enhanced. Finally, the outlet boundary condition relies on a second-order extrapolation of the velocity and mass density. The extrapolated distributions are evaluated through a regularization procedure relying on a finite difference estimation of the local stress tensor, as introduced in (Latt *et al.* 2008).

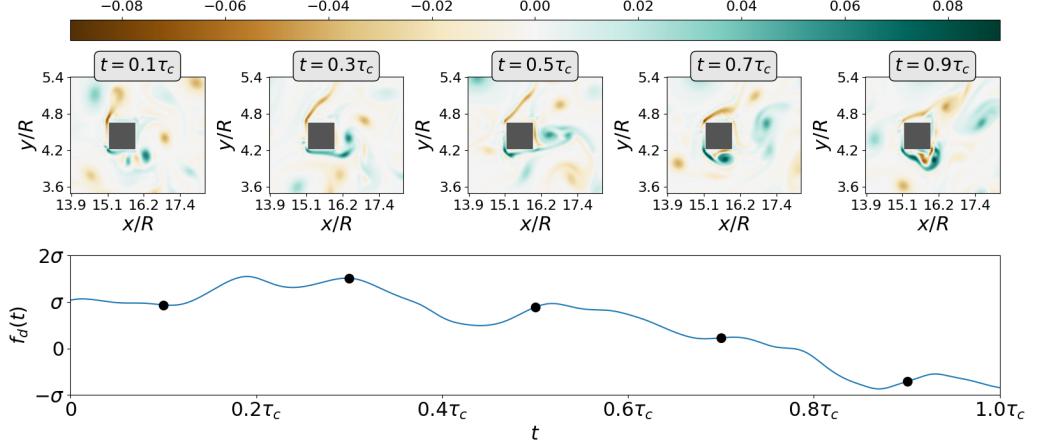


Figure 2: Snapshots of the vorticity related to typical drag fluctuations (within one standard deviation) over a time interval of length $\tau_c \simeq 4\tau_0$; τ_c will later be identified as the correlation time of the drag signal. The vorticity is given in lattice units.

2.1. The drag force

The incoming turbulent flow exerts fluctuating mechanical efforts onto the squared obstacle. The *drag* is defined as the resulting force in the streamwise x -direction. Formally

$$f_d(t) = \int_S \boldsymbol{\tau}_{x\beta}(\mathbf{x}, t) d\mathcal{S}_\beta(\mathbf{x}), \quad (2.1)$$

where \mathcal{S} is the surface of the obstacle and $\boldsymbol{\tau}$ denotes the stress tensor (see Appendix). Here, the viscous stress makes a negligible contribution to the drag. The latter therefore results mostly from pressure forces. Since the pressure on the top and bottom sides of the square applies in the normal direction, they do not contribute to the drag. As a consequence, the drag can eventually be expressed as the difference

$$f_d(t) = p_{fb}(t) - p_{base}(t) \quad (2.2)$$

between the pressure integrated over the upstream side of the obstacle or *forebody*, $p_{fb}(t)$, and the downstream side or *base* $p_{base}(t)$. Pressure fluctuations are related to the dynamics of the vorticity field. Regions of strong vorticity correspond to strong local pressure gradients, *e.g.* as demonstrated analytically with a Rankine vortex.

The typical timescale (turnover time) of drag fluctuations can be estimated from dimensional analysis as

$$\tau_0 = \frac{R}{U}, \quad (2.3)$$

where R is the size of the square and U is the averaged velocity in the channel. Fig. 2 displays the typical evolution of the vorticity field around the obstacle over a few turnover times. Because the vorticity generated along the forebody is swept away by the mean flow, the pressure field in the vicinity of the base is only slightly perturbed.

2.2. The drag as a random process

Figure 3 shows the time signal of the drag acting on the square, $f_d(t)$, over five hundred turnover times. The signal appears unpredictable in details and exhibits repeated bursts of high amplitude that deviate significantly from the averaged value. Therefore, it is natural to model the drag as a (scalar) random process.

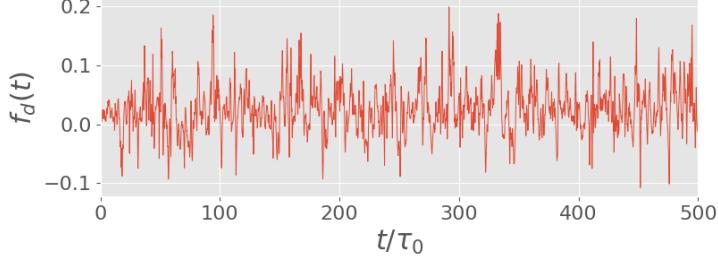


Figure 3: Temporal evolution of the drag (in lattice units) acting on the square under the action of the impinging turbulent flow. The time is normalised by the turnover time related the mean-flow velocity and the size of the obstacle, *i.e.* $\tau_0 = R/U$.

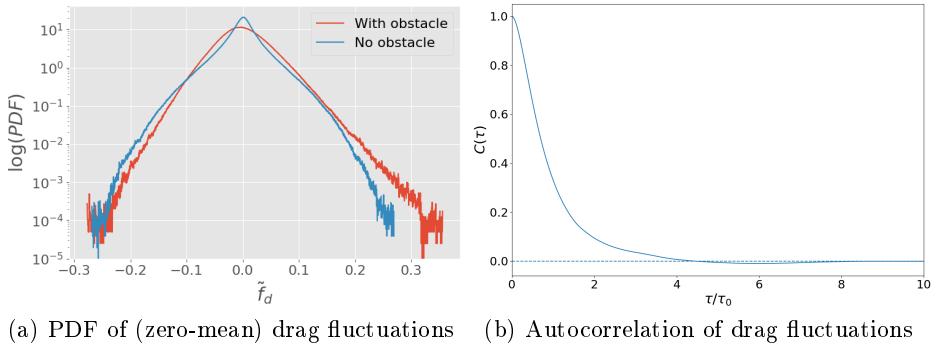


Figure 4: **(a)** PDF of (zero-mean) drag fluctuations $\tilde{f}_d \equiv f_d - \bar{f}_d$ where \bar{f}_d denotes the time-averaged value. The drag is evaluated both in the presence (red) and in the absence (blue) of the obstacle. **(b)** Autocorrelation function of the drag defined as $C(\tau) = \tilde{f}_d(t + \tau)\tilde{f}_d(t) / \overline{\tilde{f}_d^2}$. The correlation time $\tau_c \simeq 4\tau_0$ is defined by $C(\tau_c) = 0$.

Drag fluctuations have been sampled along a simulation of duration $T_{tot} = 4 \times 10^6 \tau_0$. This long simulation will be referred to as the *control run* in the following. It has been made possible by the relative simplicity of the investigated flow and the computational efficiency of the lattice Boltzmann method. The Probability Density Function (PDF) of drag fluctuations is shown in Fig. 4a. It deviates from a normal law and shows an exponential tail for large positive fluctuations, *i.e.* $\mathbb{P}(f_d) \propto e^{-\lambda f_d}$. Fig. 4a also displays the PDF of drag fluctuations acting on a control surface corresponding to the periphery of the obstacle but in the absence of the obstacle. In that case, the PDF is quasi-symmetric and does not display exponential tails. This shows that the asymmetry of the PDF and the development of a positive exponential tail are closely related to the no-slip condition on the obstacle boundary.

Lastly, the autocorrelation function of the drag $C(\tau)$ is shown in Fig. 4b. It is found that drag fluctuations are correlated over a time interval $\tau_c \simeq 4\tau_0$. One can then argue that the drag loses its memory over a time scale corresponding to the sweeping of a few eddies past the obstacle. This observation is important for the application of rare-event algorithms as it will be discussed in section 4. In the following, τ_c will be referred to as the *correlation time* of the drag process. The ratio τ_0/τ_c may be viewed as a Strouhal number. The value $St = 0.25$ is consistent with common observations for flows past blunt structures at comparable Reynolds numbers.

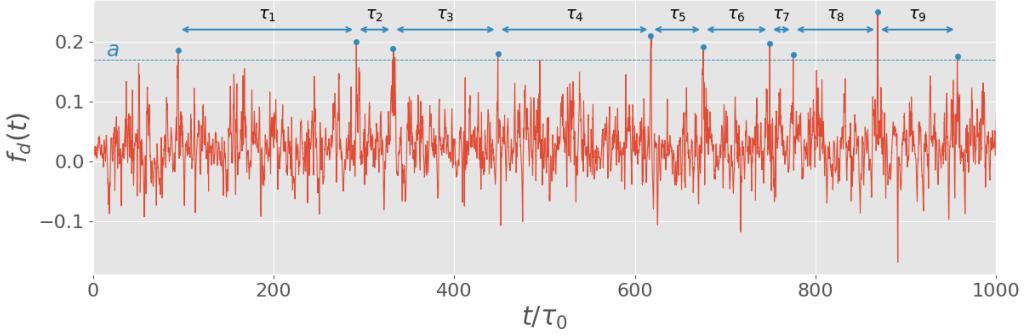


Figure 5: The *return time* $r(a)$ is the averaged waiting time between the occurrence of peak fluctuations of amplitude larger than a . One observes that $r(a) \gg \tau_c$ (correlation time) if a is sufficiently large. The selected peak fluctuations are therefore well separated.

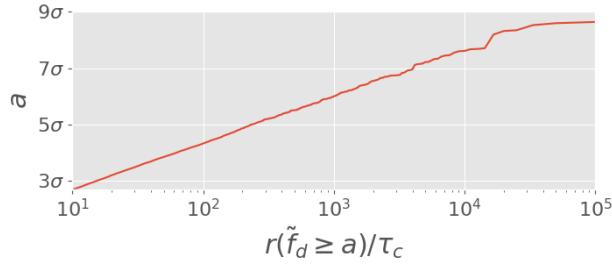


Figure 6: Amplitude of drag fluctuations as a function the corresponding return time. \tilde{f}_d denotes the drag with zero mean, *i.e.* $\tilde{f}_d = f_d - \bar{f}_d$.

3. Extreme fluctuations of the drag by means of direct sampling

The phenomenology of extreme fluctuations of the drag is first investigated through brute-force direct sampling applied to the control run. Direct sampling is here used as opposed to approaches involving rare-events algorithms discussed in section 4. It will provide a trustworthy baseline for the validation of rare-events algorithms.

The waiting times τ are defined as the time between two consecutive occurrences of peak fluctuations with amplitude $f_d \geq a$, as illustrated in Fig. 5. The mixing time τ_m is the time needed for the dynamics to loose the memory of its initial conditions. As soon as the typical waiting times are much larger than the mixing time τ_m , the occurrences of such events follow a Poisson process and the distribution of the waiting times is exponential, *i.e.* $P(\tau) = \lambda(a) \exp(-\lambda(a)\tau)$ where $r(a) = 1/\lambda(a)$ is the averaged waiting time (Lestang *et al.* 2018); $r(a)$ is called the *return time* of the level a . For systems without multi-stability, it is common for the mixing time τ_m to be of the order of the correlation time τ_c .

How rare is a fluctuation a is quantified by its return time $r(a)$. We can define extreme drag fluctuations as *rare events* in the sense that the return time is much larger than the correlation time, *i.e.* $r(a) \gg \tau_c$. If one assumes that $r(a) = t(a) / \mathbb{P}(f_d \geq a)$ where the time scale $t(a)$ is of order τ_c and varies much more slowly with a than $\mathbb{P}(f_d \geq a)$, one might expect that

$$r(a) \underset{a \rightarrow \infty}{\propto} \exp(-la) \quad (3.1)$$

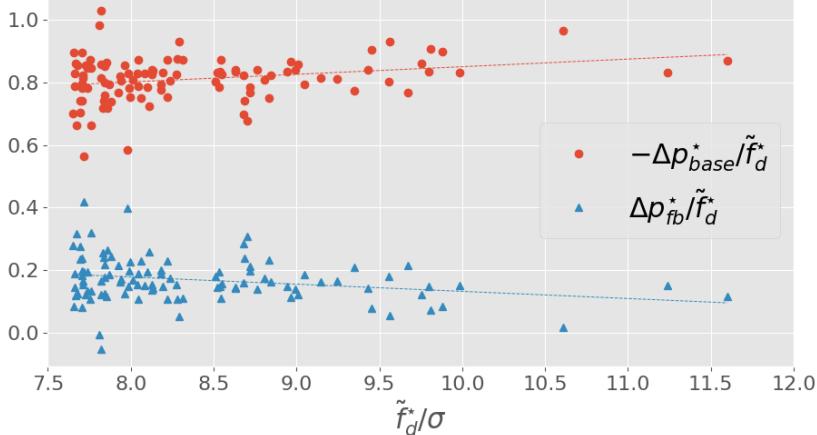


Figure 7: Relative contributions of the forebody and base pressure variations to extreme amplitudes of the drag. An extreme event corresponds to an amplitude \tilde{f}_d^* and a unique pair $(\tilde{p}_{base}^*, \tilde{p}_{fb}^*)$. As the amplitude of the drag increases, the relative variation of the base pressure also increases.

where l is the rate describing the positive tail of the PDF of the drag (shown in Fig. 4). Fig. 6 shows the evolution of the return time $r(a)$ with the amplitude of fluctuation a , computed from direct sampling of the drag signal $f_d(t)$ (Lestang *et al.* 2018). Consistently, it is found that the return time $r(a)$ is well approximated by an exponential for large levels a . Let us also point out some deviation from the exponential law at the largest levels, which are probably the consequence of under-sampling.

3.1. Extracting extreme drag fluctuations from a very long timeseries

We have extracted the fluctuations of the drag with a return time $r(a)$ greater than $10^4\tau_c$ from the control time-series $\{f_d(t)\}_{0 \leq t \leq T_{tot}}$. This set will be considered as representative of *extreme events* in the upcoming study. The choice of this particular threshold has been driven by the need to collect enough events with large amplitude and possibly identify generic features. According to Fig. 6, the related amplitude a is found equal to 7.6σ with σ being the standard deviation of the drag process. Precisely, 104 independent fluctuations with $f_d(t) \geq 7.6\sigma$ have been identified. Each fluctuation is characterized by its maximal value, f_d^* , and the time, t^* , at which this maximum is reached. In the following, the phenomenology of extreme drag fluctuations will be examined on the basis of this set of events.

3.2. Instantaneous drag

3.2.1. Contribution of forebody and base pressure fluctuations to the overall drag fluctuation

In section 2, it was pointed out that typical drag fluctuations originate mostly from the variation of the forebody pressure, *i.e.* from the upstream turbulent flow. We shall see that the situation is different in the case of extreme drag fluctuations.

Let (t^*, f_d^*) refer to an extreme-drag event. The (zero-mean) fluctuation $\tilde{f}_d^* = f_d^* - \bar{f}_d$ can be decomposed into

$$\tilde{f}_d^* = \Delta p_{fb}^* - \Delta p_{base}^* \quad (3.2)$$

where Δp_{fb}^* and Δp_{base}^* denote the variations of the forebody and base pressure, respec-

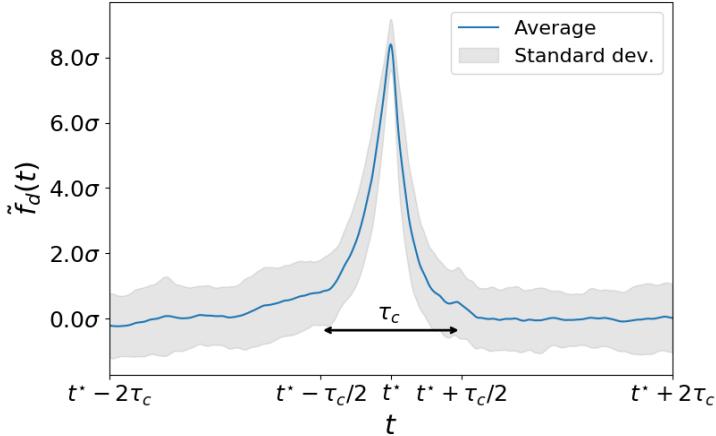


Figure 8: Ensemble average of drag signals centred around extreme fluctuations occurring at $t = t^*$. The blue line shows the mean profile whereas the shaded area indicates variations (around the mean profile) within one standard deviation. Extreme-drag events exhibit a typical lifetime of one correlation time τ_c . The profile is slightly skewed indicating that the step up is slower than the return to typical values.

tively. Fig. 7 displays the relative contributions $\Delta p_{fb}^*/\tilde{f}_d^*$ and $-\Delta p_{base}^*/\tilde{f}_d^*$ to the overall drag fluctuation \tilde{f}_d^* . It is found that the base pressure variation contributes typically to 80% of the overall drag fluctuation. Therefore, extreme amplitudes of the drag are dominated by the variation of the pressure in the vicinity of the base of the obstacle, *i.e.* downstream of the obstacle. Furthermore, Fig. 7 suggests that the larger the fluctuation, the more important is the relative contribution of the base pressure.

3.2.2. Fluid dynamics related to extreme drag fluctuations

The focus is now on the flow scenarii that yield extreme values of the drag. Fig. 8 displays the mean profile (in time) of the drag signal around extreme events. A peaked profile is observed with a width roughly corresponding to one correlation time τ_c . This shows that the duration of extreme events corresponds typically to the sweeping time of the flow past the obstacle. Interestingly, the profile is also slightly skewed indicating that the step up of the drag is slower than the return to typical values past the peak value. This is reminiscent of time-irreversibility in turbulent dynamics. To better understand the flow scenarii leading to these events, the vorticity fields around the obstacle are now examined.

Fig. 9 displays the vorticity field (in lattice units) around the obstacle for the highest amplitudes of the drag during the control run. In each case, an intense vortical structure is visible near the base of the obstacle. The vorticity level of this structure is typically twice the amplitude of typical vorticity fluctuations observed in Fig. 2. The formation of this vortex originates from an intense negative (or positive) vorticity layer at the top (or bottom) boundary of the obstacle. This high vorticity is responsible for a significant pressure drop at the base of the obstacle and therefore a strong drag. In contrast, nothing special happens near the forebody of the obstacle during extreme-drag events.

The high pressure drop near the base of the obstacle appears to be closely related to the presence of a strong vortex blocked against the base. As illustrated in Fig. 10, this blockage is enforced by the presence of opposite vorticity in the near wake, which holds the vortex against the base of the obstacle and prevents it from being swept away for a

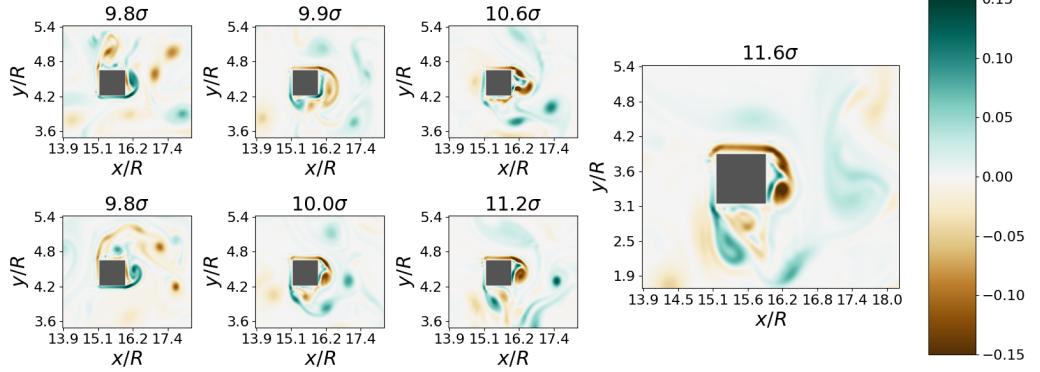


Figure 9: Vorticity field (in lattice units) around the obstacle at $t = t^*$ for the highest drag amplitudes recorded in the control run.

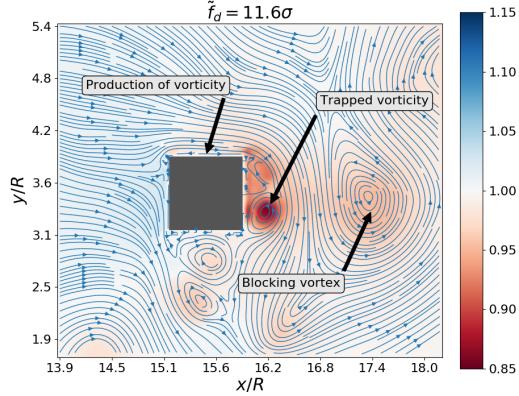


Figure 10: Pressure field (in lattice units) and velocity streamlines at $t = t^*$. In the near wake of the square, a (blocking) vortex blocks an intense vortex against the base of the obstacle.

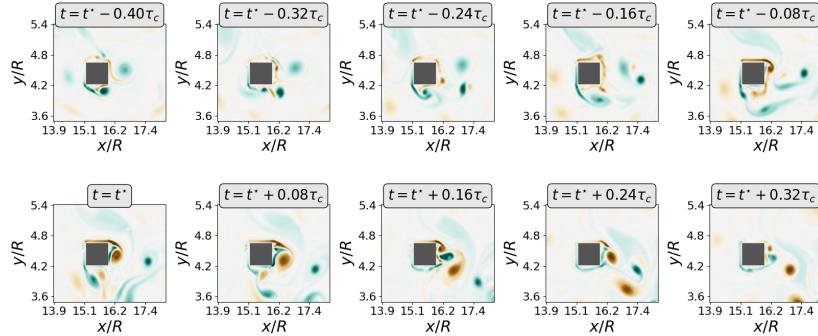


Figure 11: Snapshots of the vorticity field (in lattice units) around $t = t^*$.

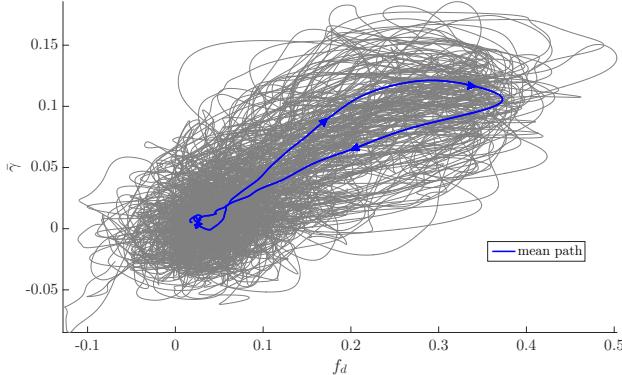


Figure 12: Evolution of the (integrated) shear along the top or bottom sides of the obstacle as a function of the drag for $t^* - 2\tau_c \leq t \leq t^* + 2\tau_c$. Each trajectory corresponds to a single event. The blue line is the mean path averaged over the set of extreme events sampled in the control run.

while. This scenario is better evidenced by Fig. 11, where the time history of the vorticity field around $t = t^*$ for the same event is shown. Before the occurrence of the extreme event, positive vorticity originating from the bottom boundary layer develops in the near wake of the square. This positive vorticity prevents the shedding of negative vorticity and enforces the development of a intense vortex against the base of the square. As the blocking vortex is in turn advected downstream, the vortex against the base is released. Consistently, one can argue that the typical duration of this scenario is related to the sweeping time of the flow past the obstacle, and is therefore of the order of τ_c . This is in full agreement with the typical duration obtained from statistical consideration on the mean profile of large-drag fluctuations in Fig. 8. This scenario is generic and has been observed for most extreme events sampled in the control run.

Since the occurrence of large drag amplitudes arises from the production of vorticity along the top or bottom side of the square, it is proposed to characterize the dynamics of extreme events by their trajectory in the parameter space $(f_d(t), \bar{\gamma}(t))$ where $\bar{\gamma}(t)$ is the averaged shear along the top or bottom boundary of the square:

$$\bar{\gamma} = \frac{1}{R} \int_{\mathcal{S}_{\parallel}} \frac{\partial u(\mathbf{x})}{\partial y} d\mathbf{x}, \quad (3.3)$$

where R denotes the size of the square, u is the streamwise component of the velocity field and \mathcal{S}_{\parallel} is the surface of either the top or the bottom boundary. Fig. 12 shows $\bar{\gamma}(t)$ as a function of the instantaneous drag $f_d(t)$ for $t^* - 2\tau_c \leq t \leq t^* + 2\tau_c$ for the 104 sampled extreme events. Before and after the extremal fluctuation, *i.e.* for $t^* - 2\tau_c \leq t \leq t^* - \tau_c$ and $t^* + \tau_c \leq t \leq t^* + 2\tau_c$, paths wander in a region related to typical values of both $\bar{\gamma}$ and f_d . On the contrary, the drag abruptly varies for $t^* - \tau_c \leq t \leq t^* + \tau_c$ near the extremal amplitude. These excursions always go clockwise, that is, $\bar{\gamma}$ attains its maximum value before f_d does. This is consistent with an increase of $\bar{\gamma}$ acting as a precursor for extreme drag fluctuations. In this representation, we also observe that the path related to the increase of the drag is longer than the path related to the return to typical values.

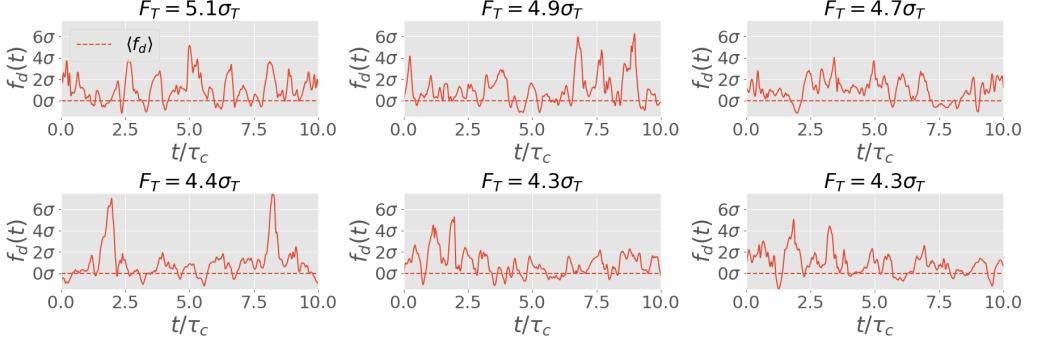


Figure 13: Instantaneous drag signals $f_d(t)$ corresponding to the highest fluctuations of the averaged drag F_T with a time window $T = 10\tau_c$; σ and σ_T denote the standard deviations of the instantaneous and averaged drag, respectively.

3.3. Extreme fluctuations of the time-averaged drag

We discussed previously the phenomenology of extreme fluctuations of the *instantaneous* drag, and identified the sweeping time of the flow past the obstacle as the characteristic lifetime of these events. In applications, this duration may be much smaller than the response time of the material structure subject to these fluctuations, justifying a practical interest in the averaged (in time) drag force. Therefore, a relevant observable is the *time-averaged* drag defined as

$$F_T(t) = \frac{1}{T} \int_t^{t+T} f_d(t) dt, \quad (3.4)$$

where $f_d(t)$ denotes the instantaneous drag and T is the investigated timescale (response time). Relevant values of T are application dependant. We shall consider $T = 10\tau_c$ in the following.

During a time interval $[t; t + T]$, a fluctuation of $F_T(t)$ may be roughly viewed as the overall contribution of T/τ_c independent fluctuations of the instantaneous drag f_d . It is thus legitimate to ask whether a large value of the averaged drag results from a single outstanding fluctuation of the instantaneous drag (case (1)), or from an unusual succession of moderate positive fluctuations (case (2)). In the same way as in section 3.1, one can identify extreme fluctuations of F_T exceeding some fixed threshold a , and sample a set of extreme events. By taking $a = 5.2\sigma_T$ with σ_T being the standard deviation of F_T , 84 independent extreme events have been selected.

Fig. 13 displays the time-series $\{f_d(t)\}_{t^* \leq t \leq t^* + T}$ for several extreme fluctuations of F_T occurring at $t = t^*$. It is found that extreme fluctuations of the time-averaged drag can neither be reduced to case (1) nor case (2). Indeed, both cases are equally featured in Fig. 13. Very large value of the averaged-drag results from either a very large fluctuation, or a significant succession of moderate (positive) fluctuations of the instantaneous drag. This observation can be related to the exponential shape of the tail of the PDF describing extreme positive drag fluctuations. Indeed, let X be a random variable with a PDF $\mathbb{P}(X)$ and a standard deviation σ_X . Considering an extreme positive value of $S_N = \sum_{n=1}^N X_n$, the probability p_1 (resp. p_2) of case (1) (resp. case (2)) writes

$$p_1 \left(\sum_{n=1}^N X_n = S_N \right) \propto \mathbb{P}(S_N/N)^N \quad \text{and} \quad p_2 \left(\sum_{n=1}^N X_n = S_N \right) \propto \mathbb{P}(S_N) \quad (3.5)$$

If \mathbb{P} has an exponential positive tail, *i.e.* $\mathbb{P}(X = x) \underset{x \gg \sigma_X}{\propto} e^{-\lambda x}$, both cases (1) and (2) have equivalent probabilities provided that the average $a = S_N/N$ is very large:

$$\frac{p_2}{p_1} \underset{a \rightarrow \infty}{\rightarrow} (e^{\lambda a})^N e^{-\lambda a N} = 1 \quad (3.6)$$

The observation is therefore well supported theoretically. See (Lestang *et al.* 2019) for further details.

4. Rare-event algorithms

In the limit of very rare events and complex dynamics such as turbulent flows in industrial or environmental domains, the computational cost of direct sampling becomes prohibitive. According to Eq. (3.1), the return time of fluctuations $f_d \geq a$ scales like $r(a) \propto e^{la}$. As a consequence, the computational cost required to sample events of amplitude $f_d \geq a$ through brute-force sampling diverges exponentially. The motivation behind rare-events algorithms is to sample extreme fluctuations for a computational cost much lower than their related return times. For this purpose, the Adaptive Multilevel Splitting (AMS) and Giardina-Kurchan-Tailleur-Lecomte (GKTL) algorithms are considered in the following. Both algorithms rely on an ensemble of N trajectories $\{\mathbf{x}_n(t)\}_{0 \leq t \leq T_a}$ with $n = 1 \dots N$. Each $\{\mathbf{x}_n(t)\}_{0 \leq t \leq T_a}$ refers to a trajectory (of duration T_a) of the system in phase space. In our case, the dynamical system under investigation is the flow described in section 2. Trajectories are replicated or discarded in the ensemble according to some *selection rules* designed to sample rare events of some observable, *e.g.* the drag, with a higher frequency. These selection rules are such that the introduced statistical bias is known at each iteration of the algorithm. The generated trajectories are associated with a weight, from which one is able to compute the probability and expectation values of observables.

4.1. Extreme instantaneous drag forces with the Adaptive Multilevel Splitting algorithm

The Adaptive Multilevel Splitting algorithm (Cérou & Guyader 2007) builds on previous ideas of splitting algorithms (Kahn & Harris 1951). The sampling of a rare event is made easier by splitting the dynamical path of the system (in phase space) into a sequence of shorter excursions (Glasserman *et al.* 1999; Rolland & Simonnet 2015). The related selection-mutation step of this algorithm relies on the definition of a *score function* $\xi(\mathbf{x}(t), t)$. The operating principle is sketched in Fig. 14. Iterations of the algorithm consist in discarding the trajectories with the lowest maxima of the score function during the time interval $0 \leq t \leq T_a$. Discarded trajectories are re-sampled according to the remaining trajectories. See (Cérou *et al.* 2019) for a review of the AMS algorithm, its history and applications. For recent applications of the AMS in the context of fluid turbulence, see (Bouchet *et al.* 2019; Rolland 2018). Let us mention that we use a specific form of the AMS, in which trajectories have a fixed duration T_a . It can however be formulated in terms of the “classical AMS” (Lestang *et al.* 2018). The following results are not limited to this particular variant.

4.1.1. Illustration of the AMS on a simple case: the Ornstein-Uhlenbeck process

Fluid dynamics is temporarily left aside and a one-dimensional Ornstein-Uhlenbeck process is considered:

$$\dot{x} = -x + \eta(t), \quad (4.1)$$

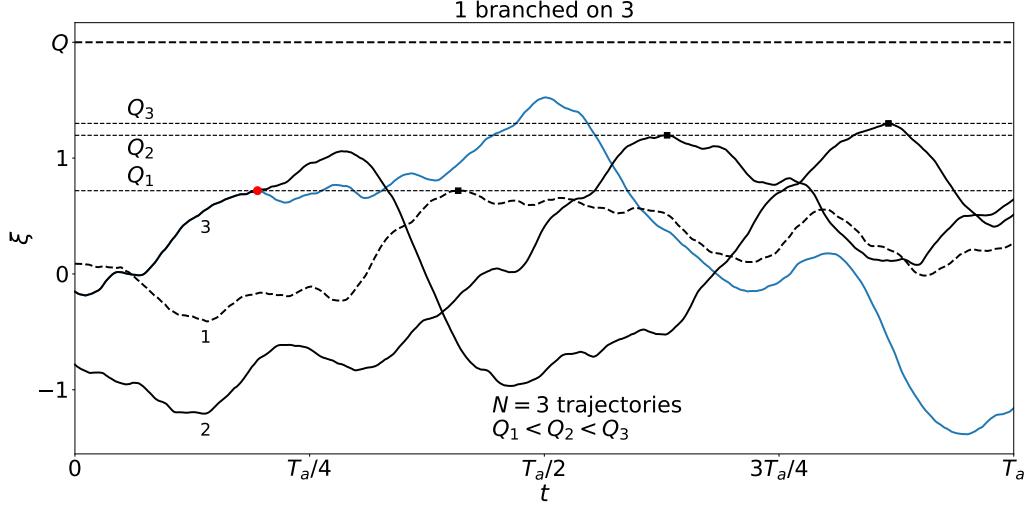


Figure 14: Selection-mutation procedure of the AMS algorithm with $N = 3$. Lines 1, 2 and 3 represent the evolution of the score function for the current ensemble of trajectories. On the basis of their respective maximum: Q_1 , Q_2 and Q_3 , the trajectory with the lowest maximum is discarded in the ensemble (dashed line). Among the two remaining trajectories, trajectory 3 is chosen randomly and copied until it reaches the value Q_1 . It is then simulated from the branching point to the final time T_a . In case of deterministic dynamics, a small perturbation is introduced at the branching to separate the trajectories. Depending on the application, this re-sampling procedure can be iterated J times or until all trajectories do exceed a fixed threshold Q .

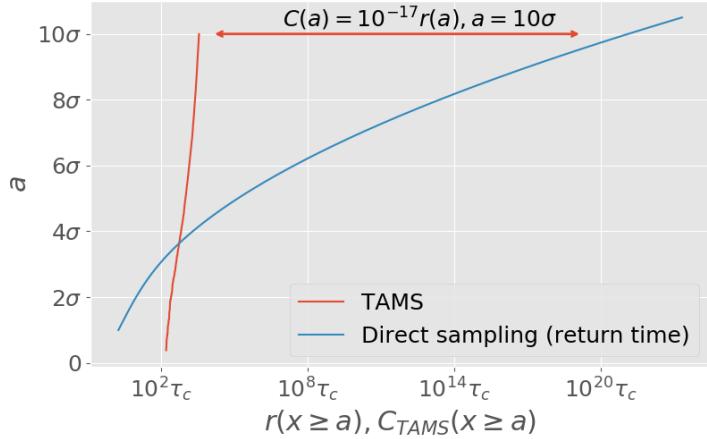


Figure 15: Efficiency of the TAMS algorithm with respect to direct sampling in the case of an Ornstein-Uhlenbeck process (Lestang *et al.* 2018). The red line represents the evolution of the maximum obtained from re-sampled trajectories as a function of the computational cost C_{TAMS} . The blue line is the analytical solution for the return time of amplitude a .

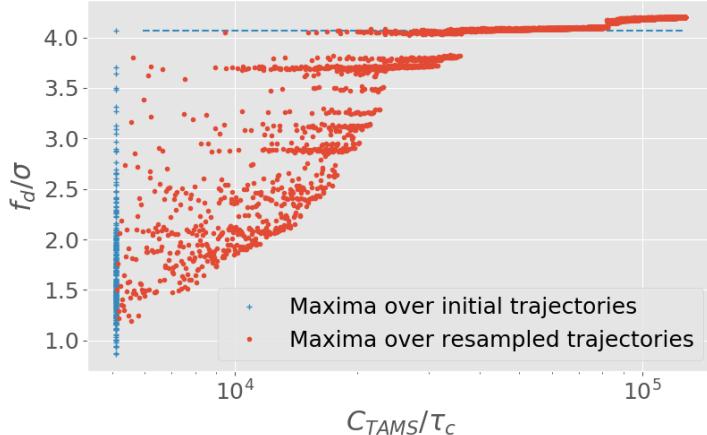


Figure 16: Maximum of the instantaneous drag throughout the re-sampled trajectories as a function of the corresponding computational cost C_{TAMS} . The AMS is unable to efficiently sample rare trajectories associated to drag fluctuations higher than the largest fluctuation already captured in the initial ensemble.

where η is a Gaussian noise with $\langle \eta(t)\eta(t-t') \rangle = \delta(t-t')$. This basic stochastic process will allow us to highlight differences with fluid dynamics.

The AMS is applied to a set of $N = 32$ trajectories $\{x_n(t)\}_{0 \leq t \leq T_a}$ with $T_a = 5\tau_c$. Let us note that the correlation time is $\tau_c = 1$ for the process defined by Eq. (4.1). Our objective is to sample fluctuations $x \geq a$ with a being very large compared to the typical values of x . The score function is simply $x(t)$ and a single trajectory is re-sampled at each iteration. The computational cost of the algorithm after J iterations is therefore related to the simulation of the N initial trajectories and the re-sampling of J trajectories. Fig. 15 compares the computational cost of the AMS algorithm with that of a direct sampling. In the latter, the typical computational cost is simply the return time $r(a)$. One can see that the successive re-samplings of the AMS algorithm lead rapidly to trajectories exhibiting extreme fluctuations. For large a , the computational cost is many orders of magnitude lower than that obtained by direct sampling.

Undoubtedly, the Ornstein-Uhlenbeck process has oversimplified dynamics to showcase the efficiency of the AMS algorithm. The state space is one-dimensional and the choice of the score function is straightforward: It is x itself. In addition, the noise term in Eq. (4.1) has no correlation in time, which implies that newly generated trajectories quickly separate from their parents. Such favorable features do no *a priori* persist in the case of fluid dynamics.

4.1.2. The AMS for extreme drag fluctuations

Our aim is now to use the AMS algorithm to sample flow evolutions that exhibit extreme fluctuations of the drag f_d acting on the square obstacle. In contrast with the Ornstein-Uhlenbeck process, the phase space is here highly-dimensional with intrinsically chaotic dynamics. In that situation, the choice of the score function $\xi(\mathbf{x}(t), t)$ is no longer straightforward. However a natural simple choice is $\xi(\mathbf{x}(t), t) = f_d(t)$, for which we opt for in the following. Let us also stress that the dynamics is here deterministic. Therefore, randomness must be artificially introduced for the re-sampling procedure to be effective, *i.e.* to generate new trajectories associated to larger drag fluctuations. In practice, a small perturbation in the state of the flow has been introduced at the branching points

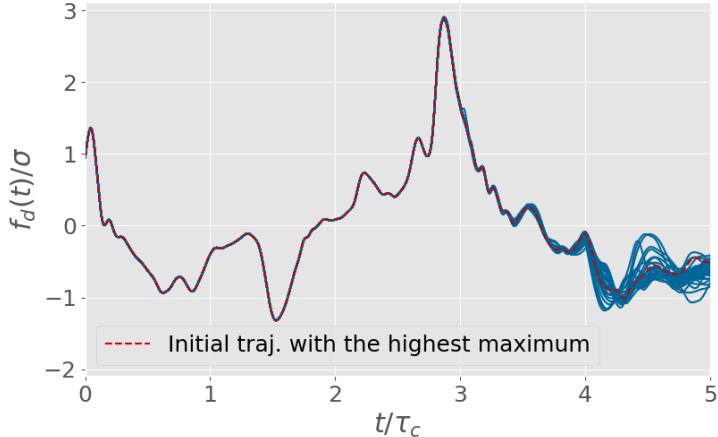


Figure 17: Ensemble of $N = 32$ trajectories after 181 iterations of the selection-mutation procedure. In this experiment, the AMS algorithm is used with the instantaneous drag $f_d(t)$ as score function. Each trajectory has a duration $T_a = 5\tau_c$ (correlation time of the drag). The AMS fails to efficiently sample rare trajectories. All trajectories are eventually re-sampled from the same trajectory displaying the highest maximum in the initial ensemble.

(see Fig. 14 and (Wouters & Bouchet 2016)). Due to the chaotic behavior of the flow, re-sampled trajectories then separate from their parents. We assume that this perturbation at branching times has a negligible impact on the statistical properties of the sampled rare events. The validity of this assumption has been tested by performing a long simulation of the dynamics and by regularly perturbing the flow. It was checked that the obtained statistics of the drag were consistent with the statistics computed from the (unperturbed) control simulation described in section 2.2.

The AMS algorithm is applied to $N = 256$ trajectories of duration $T_a = 20\tau_c$. Fig. 16 displays the maximum drag achieved by the re-sampled trajectories as a function of the computational cost, similarly to Fig. 15. The distribution of the maximal drag for the *initial* trajectories is also shown. After a few iterations, the trajectories with the lowest maxima of the score function are discarded and new trajectories with higher maxima are re-sampled. Figure 16 shows the evolution of the maxima of the re-sampled trajectories as the AMS is iterated, *i.e.* as a function of the computational cost. The re-sampled trajectories never exceed the amplitude of the highest maximum already attained in the initial set of trajectories. Note that the red points in figure 16 are analogous to the red line in figure 15 for the OU process.

Fig. 17 displays the ensemble of trajectories after many iterations of the AMS resampling procedure. Let us mention that for the sake of simplicity, the trajectory of the system is here abusively associated with the time evolution of the observed drag. Interestingly, all trajectories ultimately originate from one unique initial trajectory. This feature may be explained as follows. It takes a certain time τ_L (called Lyapunov timescale) before a re-sampled trajectory separates from its parent. In our situation, this “memory effect” originates from the fact that the score function is of dimension much smaller (here 1) than the dimension of the phase space (very large for this fluid dynamics problem). We observe in Fig. 17 that the duration of extreme drag fluctuations, τ_c , is much shorter than the Lyapunov’s timescale τ_L . As a consequence, the re-sampling of a trajectory branched close to $t = t^*$ (when the maximum drag occurs) cannot lead to larger values

at $t^* \leq t \leq t^* + \tau_L$. For $t - t^* > \tau_L$, the drag process has lost the memory of the drag fluctuations on which the re-sampling was based and therefore the probability of observing a new extreme fluctuation is also very low.

The difference between the typical duration of drag fluctuations τ_c and the Lyapunov timescale τ_L may be heuristically associated with the *turbulence rate* of the flow. As discussed previously, the duration of extreme fluctuations of the drag are closely related to the sweeping time of the flow past the obstacle, and consequently to the mean-flow velocity U . On the contrary, the Lyapunov timescale is rather associated with the intrinsic evolution of turbulent fluctuations in the reference frame of the mean flow, *i.e.* with the fluctuating velocity u_{rms} . The ratio u_{rms}/U (turbulence rate) is much lower than one in our case of grid-generated turbulence, which implies that $\tau_L \gtrsim \tau_c$.

In summary, a straightforward application of the AMS algorithm with the score function being the drag itself does not allow us to efficiently sample extreme fluctuations. Retrospectively, there is *a priori* no reason for the drag itself to be a good indicator of how likely next drag fluctuations will develop. On the other hand, finding a better score function would require a precise qualitative understanding of the flow scenario leading to extreme events. A possible line of improvement might be to explicitly account for “the correlation time of the observable being smaller than the Lyapunov time of the system” by re-sampling the trajectories at a Lyapunov time before the maximum is reached. That way, new events associated to higher fluctuations may have the opportunity to develop. However, such modification of the procedure must be implemented with care especially in order to preserve the mathematical properties of the algorithm, *e.g.* the evaluation of the statistical bias .

4.2. Importance sampling of extreme time-averaged drag forces with the Giardina-Kurchan-Tailleur-Lecomte algorithm

The sampling of extreme fluctuations of the time-averaged drag F_T is now examined. The AMS algorithm could be used in the same way as before by taking the time-averaged observable itself as the score function. However, this would lead to similar unsatisfactory results. For a time-averaged observable, an alternative approach is provided by the Giardina-Kurchan-Tailleur-Lecomte (GKTL) algorithm (Giardinà *et al.* 2006; Tailleur & Kurchan 2007; Giardinà *et al.* 2011). Similarly to the AMS algorithm, the GKTL algorithm relies on the simulation of an ensemble of trajectories. At regular time intervals, some elements of the ensemble are killed and others are cloned according to a weight that depends on the history of the element itself. The weights are chosen so that, after several iterations of the algorithm, the trajectories in the ensemble are distributed according to a biased probability distribution that favors trajectories related to large values of the time average of the observable. The GKTL algorithm belongs to a family of algorithms called “go with the winners” (Aldous & Vazirani 1994; Grassberger 2002). Similar ideas have already been applied in a wide range of fields such as polymer physics (Grassberger *et al.* 1998), out of equilibrium statistical physics (Nemoto *et al.* 2017b), computer science (Aldous & Vazirani 1994), dynamical systems (Tailleur & Kurchan 2007), quantum mechanics (Kosztin *et al.* 1996). The application of a go-with-the-winners approach to the computation of large deviations in non-equilibrium systems has first been proposed in 2006 (Giardinà *et al.* 2006). Over the last ten years, it has been successfully applied to investigate rare events in both stochastic (Giardinà *et al.* 2006; Lecomte & Tailleur 2007; Garrahan *et al.* 2007) and deterministic systems (Giardinà *et al.* 2006; Tailleur & Kurchan 2007).

4.2.1. The GKTL algorithm

The GKTL algorithm is based on the simulation of an ensemble of N trajectories $\{\mathbf{x}_n(t)\}_{0 \leq t \leq T_a}$ with $n = 1 \dots N$ starting from independent random initial conditions. Let us consider a real-valued observable of interest $A(\mathbf{x}(t))$, *e.g.* the drag $f_d(t)$, and introduce a cloning period τ . At time instants $t_i = i\tau$ with $i = 1, 2, \dots, T_a/\tau$ (T_a is a multiple of τ) a weight W_n^i is assigned to each trajectory. This weight is defined ($t_0 = 0$) by

$$W_n^i = \frac{e^{k \int_{t_{i-1}}^{t_i} A(\mathbf{x}_n(t)) dt}}{R_i} \quad \text{with the normalisation factor} \quad R_i = \frac{1}{N} \sum_{n=1}^N e^{k \int_{t_{i-1}}^{t_i} A(\mathbf{x}_n(t)) dt} \quad (4.2)$$

so that $\sum_{n=1}^N W_n^i = N$. The weights $\{W_n^i\}_{n=1 \dots N}$ determine how many copies of each trajectory are made at time $t = t_i$. The parameter k characterizes the amplitude of the statistical bias involved in the algorithm (see Fig. 18). For more information about the practical implementation of the algorithm, the interested reader can refer to (Brewer *et al.* 2018; Lestang 2018). The application of this re-sampling at each step t_i eventually leads to a biased sampling in the trajectory space; the trajectories corresponding to extreme values of $\int_0^{T_a} A(\mathbf{x}_n(t)) dt$ have a higher probability. The sampled biased distribution writes

$$\mathbb{P}_k \left(\{\mathbf{X}(t)\}_{0 \leq t \leq T_a} = \{\mathbf{x}(t)\}_{0 \leq t \leq T_a} \right) \underset{N \rightarrow \infty}{\sim} \frac{e^{k \int_0^{T_a} A(\mathbf{x}(t)) dt}}{Z(k, T_a)} \mathbb{P}_0 \left(\{\mathbf{X}(t)\}_{0 \leq t \leq T_a} = \{\mathbf{x}(t)\}_{0 \leq t \leq T_a} \right), \quad (4.3)$$

where $\mathbb{P}_0 \left(\{\mathbf{X}(t)\}_{0 \leq t \leq T_a} = \{\mathbf{x}(t)\}_{0 \leq t \leq T_a} \right)$ refers formally to the probability of observing the trajectory $\{\mathbf{x}(t)\}_{0 \leq t \leq T_a}$. The normalisation factor is given by $Z(k, T_a) = \prod_{i=1}^{T_a/\tau} R_i$. One can mention that

$$Z(k, T_a) \underset{N \rightarrow \infty}{\sim} \mathbb{E}_0 \left[e^{k \int_0^{T_a} A(\mathbf{X}(t)) dt} \right], \quad (4.4)$$

with \mathbb{E}_0 being the expectation value with respect to the distribution \mathbb{P}_0 . This result relies on the *mean-field approximation*

$$R_i = \frac{1}{N} \sum_{n=1}^N e^{k \int_{t_{i-1}}^{t_i} A(\mathbf{x}_n(t)) dt} \underset{N \rightarrow \infty}{\sim} Z(k, t_i) = \mathbb{E}_i \left[e^{k \int_{t_{i-1}}^{t_i} A(\mathbf{X}(t)) dt} \right], \quad (4.5)$$

where $\mathbb{E}_i[\cdot]$ denotes the expectation value with respect to the biased distribution $\mathbb{P}_k^{(i)}$ obtained after i cloning steps. The typical relative error related to this approximation can be shown to be of order $1/\sqrt{N}$ for a family of rare-event algorithms including the GKTL algorithm (Moral 2004b; Del Moral 2013). Rejected trajectories are discarded from the statistics. Eventually, an effective ensemble of N trajectories of duration T_a is obtained, distributed according to \mathbb{P}_k .

A key feature of the GKTL algorithm is that the sampling procedure does not involve any alteration of the dynamics. All trajectories in the resampled ensemble are solutions of the original dynamical. Nevertheless, it should be noted that a small random perturbation is introduced in the cloning procedure to allow clones of a same trajectory to separate. As for the AMS algorithm, it is assumed that this perturbation does not significantly affect the statistics of the sampled trajectories. Eventually, the sampled trajectories obtained with the GKTL algorithm can be used to compute the statistical properties of any observable with respect to the distribution \mathbb{P}_0 from the distribution \mathbb{P}_k by using Eq. (4.3).

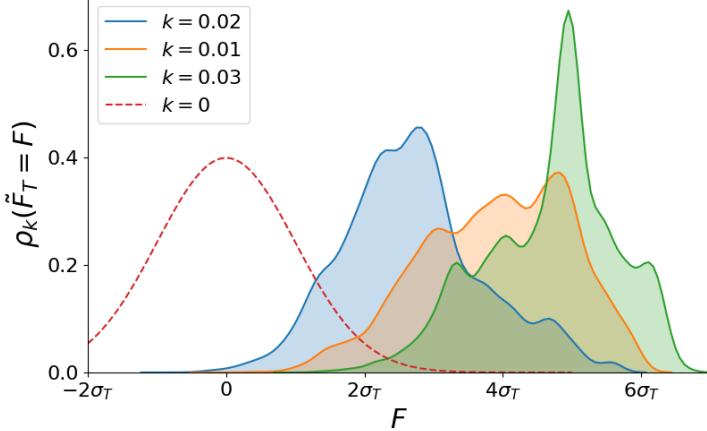


Figure 18: Rare-event sampling of the (zero-mean) time-averaged drag $\tilde{F}_T = F_T - \overline{F}_T$ with $T = 10\tau_c$; τ_c is the correlation time of the instantaneous drag. The shaded PDFs are estimated from the biased ensemble resulting from the GKTL algorithm applied to 16384 trajectories of duration $T_a = T$ with the cloning period $\tau = \tau_c/2$. The dashed line refers to the unbiased PDF of \tilde{F}_T , *i.e.* obtained from direct sampling or with $k = 0$ (no bias) in the GKTL algorithm.

4.2.2. Application of the GKTL algorithm to extreme fluctuations of the time-averaged drag

The application of the GKTL algorithm is now considered for the flow dynamics introduced in section 2. The purpose is to sample trajectories with extreme fluctuations of the *time-averaged* drag, F_T . Referring to the previous notations, $A(\mathbf{x}(t)) = f_d(t)$ is our observable of interest and the duration T_a of each trajectory corresponds to the period of averaging of the drag, *i.e.* $T_a = T$.

The computational cost C_{gklt} depends on both the duration of each trajectory and the number of trajectories: $C_{gklt} = N \times T_a$. We have taken $N = 16384$ and $T_a = 10\tau_c$ which yields $C_{gklt} \approx 1.6 \times 10^5 \tau_c$. The choice of the cloning period τ is important in practice. A cloning period too-short can result in a loss of information if the clones do not separate from their parents between two cloning steps. On the contrary, choosing $\tau \gg \tau_c$ may result in insufficient cloning steps to allow for efficient importance sampling. As a result, a safe rule of thumb is to take τ of the order of τ_c . In the following experiments, $\tau = \tau_c/2$.

Three numerical experiments corresponding to three different values of the parameter k have been carried out. Fig. 18 shows the estimate of the biased PDFs of the (zero-mean) time-averaged drag $\tilde{F}_T \equiv F_T - \overline{F}_T$, *i.e.*

$$\rho_k(F) \approx \frac{1}{N} \sum_{j=1}^N \delta(\tilde{F}_T(\{\mathbf{x}_j\}_{0 \leq t \leq T_a}) - F) \quad (4.6)$$

in addition to the unbiased PDF based on the control run. As expected, the biased PDFs obtained by applying the GKTL algorithm are centered around increasing fluctuations with the parameter k . Nevertheless, let us mention that for a fixed number of trajectories, there is necessarily an upper limit k_{max} over which the finite number of trajectories becomes detrimental to the efficiency and accuracy of the selection procedure. For $k \gtrsim k_{max}$, the re-sampling relies only on a small number of “independent trajectories” and most of the trajectories in the biased ensemble overlap. This effect is highlighted in

	σ	2σ	3σ	4σ	5σ	
$ k = 0.02 $	1594	799	155	22	0	
$ k = 0.025 $	1019	834	521	198	27	
$ k = 0.03 $	539	510	391	205	36	
$ N_{\text{brute-force}} $	2599	37	22	0.5	0.005	

Table 1: Number of fluctuations $F_T \geq a$ with $a=\sigma$, 2σ , etc. captured in the biased ensemble for $k = 0.02$, 0.025 and 0.03 . Note that trajectories that overlap for more than half of their duration are counted as only one. $N_{\text{brute-force}}$ is the number of fluctuations expected from the direct sampling of N independent realizations of F_T .

Fig. 18 where the biased PDF corresponding to $k = 0.03$ is artificially peaked. Such a finite-size effect is expected to gradually increase as the bias amplitude k is increased. An important question concerning the application of the GKTL algorithm is therefore that of the dependence of k_{\max} on the ensemble size N . This issue is addressed in recent studies (Nemoto *et al.* 2017a; Hidalgo 2018). In the present work, the order of magnitude of k_{\max} has been estimated *empirically* by evaluating the diversity of the trajectories in the biased ensemble. This diversity has been monitored at each selection step of the algorithm by calculating the proportion of trajectories sharing the same ancestor trajectory.

The efficiency of the GKTL algorithm is now assessed with respect to direct sampling. Table 1 indicates the number of trajectories that correspond to a time-averaged drag $\tilde{F}_T \geq a$ for different values of the bias parameter k . Note that trajectories that overlap for more than half of their duration are counted as one. Table 1 also indicates the expected number of fluctuations obtained by simulating N independent trajectories $\{F_T(t)\}_{0 \leq t \leq T_a}$ without resampling. We obtain that the GKTL algorithm is able to sample drag fluctuations that are far from reach by direct sampling with an equivalent computational cost.

4.2.3. Computation of return times

Statistics of extreme events can be computed from the biased ensemble by inverting Eq. (4.3). For instance, it is possible to compute return times for events unreachable by a direct approach.

In the following, the GKTL algorithm is used with $N = 1024$ trajectories of duration $T_a > T = 10\tau_c$. Each trajectory in the biased ensemble results in a timeseries of the time-averaged drag

$$F_T(t) = \int_{t-T}^t f_d(\tau) d\tau, \quad t \in [T, T_a] \quad (4.7)$$

and the return time of a fluctuation $F_T \geq a$ is given by (Lestang *et al.* 2018)

$$r(a) = -\frac{T_a - T}{\ln(1 - \mathbb{P}(F_T \geq a))}. \quad (4.8)$$

The probability $\mathbb{P}(F_T \geq a)$ can be estimated from the biased ensemble by inverting

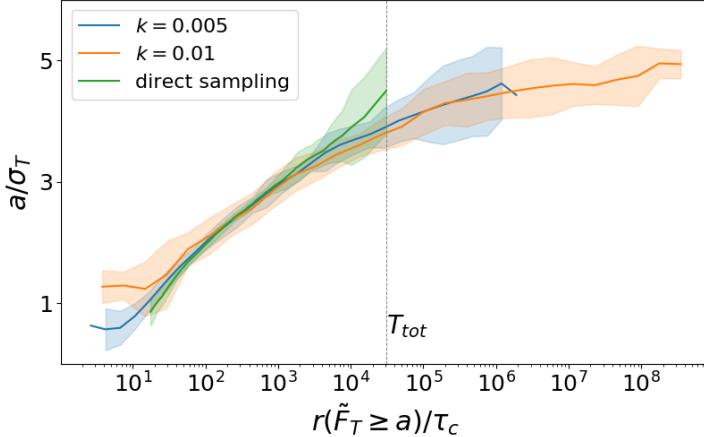


Figure 19: Return times for the time-averaged drag acting on the square obstacle. \tilde{F}_T denotes the time-averaged drag with zero mean. The blue and red lines are obtained from the biased ensemble of trajectories generated by the GKTL algorithm with $N = 1024$ and $T_a = 30\tau_c$. The green line is the return times obtained from a single timeseries of duration equal to the computational cost of both GKTL experiments. Uncertainty ranges for the GKTL estimates are computed as the standard deviation over a set of 10 independent experiments. Uncertainty ranges for the direct estimation are computed as the standard deviation over a ensemble of direct estimates resulting from 60 independent timeseries.

Eq. (4.3)

$$\mathbb{P}(f_d \geq a) \approx \frac{1}{N} \sum_{j=1}^N e^{T_a \lambda(k)} e^{k \int_0^{T_a} f_d(t) dt} s_j(a) \quad (4.9)$$

with $s_j(a) = 1$ if $\max_{T \leq t \leq T_a} [F_T^{(j)}] \geq a$ and $s_j(a) = 0$ otherwise, *i.e.* by summing the weights of the timeseries which maximum is larger than a .

Fig. 19 displays the return times for extreme fluctuations of the time-averaged drag acting on the square obstacle. The figure shows two independant estimates, obtained using different values of the bias parameter k . Note that both estimates have been computed with the same computational cost $T_{tot} = N \times T_a$. In addition, figure 19 shows an estimate computed through direct sampling with the same computational cost, *i.e.* using a timeseries of duration T_{tot} . Whilst a direct approach cannot access events with a return time greater than T_{tot} , the GKTL algorithm allows us for the computation of statistics for drag fluctuations having a return time several orders of magnitude above T_{tot} . This is obviously a major advantage of this rare-event sampling algorithm.

5. Conclusion

In this study, the application of two rare-event sampling algorithms has been assessed for the simulation of extreme mechanical efforts on structures immersed in a turbulent flow; a situation relevant to many industrial applications.

In a first part, we investigated the dynamics and statistics of extreme fluctuations of the drag acting on a square mounted in a turbulent channel flow, in two dimensions. By means of a long simulation of reference, we observed that such extreme events are caused

by the (temporary) trapping of vorticity very close to the base of the square. Extreme drag amplitudes do not persist over time since the mean flow eventually sweeps away the fluid structures responsible for this situation. The lifetime of extreme drag fluctuations is therefore of the order of the turnover time, and the corresponding drag signal is very peaked around these extreme values. Our long drag timeseries also reveal that the tails of the PDF for the drag are well described by an exponential PDF. This property is linked to the phenomenology of extremes of the time-averaged drag. Especially, we observed that such extreme values for the average do not preferentially result from a small number of very large fluctuations, or an exceptional succession of moderate fluctuations that pile up to yield an large value of the average. Such a phenomenology is observed for stochastic processes with an exponential PDF.

The application of rare event algorithms relies on the definition of a score function for the selection of trajectories. The efficiency of the algorithm depends on the choice of the score function which has to be well suited to the phenomenology of the extreme dynamics. For complex dynamics including turbulent flows, this choice is difficult because of the lack of characterisation of rare events. These latter are expected to depend on the dynamical regime (Reynolds number), the stirring mechanism and boundary conditions of the flow. In this context, the choice of the observable itself as score function is certainly a safe fall-back option, nevertheless not optimal.

On the basis of the same two-dimensional test flow, we then applied the AMS algorithm choosing the drag itself as a score function. In this case, our results illustrate that the selection-mutation procedure is unable to generate rare trajectories at a better rate than a direct sampling. This can be related to the phenomenology of extreme drag fluctuations, which lifetime is shorter than the timescale over which re-sampled trajectories separate from their parent. The GKTL algorithm has been applied to the sampling of trajectories displaying extreme fluctuations of the time-averaged drag. In this case, we showed that using GKTL leads to a significant gain with respect to a direct approach, allowing for the simulation and computation of statistics for (very) rare trajectories. The sampling of extreme time-averages is aided by the selection of trajectories displaying an exceptional succession of drag fluctuations, resulting in an extreme value of the average. It is however unclear whether the GKTL is suited to sample trajectories for which the extreme fluctuations of the time-averaged drag results from a unique, exceptionally large drag fluctuation. This issue is postponed for future investigations.

Importantly, successful usage of the AMS, GKTL or similar algorithms for complex flows relevant to industrial or environmental situations will require coping with the fact that optimal score functions are difficult to identify, if even possible. A promising direction explored in current research is to take advantage of recent advances in *learning methods* to optimise score functions beforehand.

6. Acknowledgements

The authors thank Francesco Ragone, Corentin Herbert, Charles-Edouard Bréhier and Eric Simonnet for useful discussions and suggestions on various aspects of this work. T.L and F.B acknowledge support from the European Research Council under the European Union's seventh Framework Program (FP7/2007-2013 Grant Agreement No. 616811). Simulations have been performed on the local HPC facilities at École Normale Supérieure de Lyon (PSMN) and École Centrale de Lyon (PMCS2I). The facilities at PSMN are supported by the Auvergne-Rhône-Alpes region (GRANT CPRT07-13 CIRA) and the national Equip@Meso grant (ANR-10-EQPX-29-01).

*

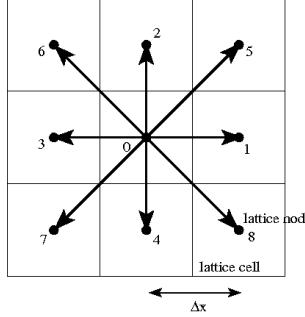


Figure 20: Sketch of the D2Q9 lattice. Particles move exactly from a lattice node towards one of its nine neighbours (including the node itself) during one time step. By definition, the lattice spacing is related to the time step by $\Delta x/\Delta t = \sqrt{3}c_s$ where c_s is interpreted as a speed of sound.

Appendix A. The Lattice Boltzmann Method

In the LB method, the fluid is viewed as a population of particles that collide, redistribute and propagate along the different links of a discrete lattice. In our two-dimensional situation, the so-called D2Q9 lattice with only nine possible velocities $\{\mathbf{c}_i\}_{i=0\dots 8}$ at each node has been adopted (see Fig. 20). Locally, the macroscopic flow variables (per unit volume) are recovered by summing over the densities of particles $\{f_i\}_{i=0\dots 8}$ moving with the different velocities, i.e.

$$\rho(\mathbf{x}, t) = \sum_i f_i(\mathbf{x}, t) \quad \text{and} \quad \rho(\mathbf{x}, t)\mathbf{u}(\mathbf{x}, t) = \sum_i f_i(\mathbf{x}, t)\mathbf{c}_i$$

for the mass density and the fluid momentum respectively. The assumption of weak compressibility (for an ideal gas) is made so that the pressure is directly proportional to the mass density: $p = c_s^2\rho$ where c_s is interpreted as a speed of sound.

The complexity of the flow emerges from the repeated application of simple rules of streaming and collision. The LBM advances the local densities of particles $f_i(\mathbf{x}, t)$ moving with velocities \mathbf{c}_i in a two-step procedure. Namely, an *exact* streaming step

$$f_i(\mathbf{x} + \mathbf{c}_i \Delta t, t + \Delta t) = f_i^{\text{out}}(\mathbf{x}, t)$$

during which particles move with their own velocity to a neighbouring node, and an instantaneous collision step

$$f_i^{\text{out}}(\mathbf{x}, t) = -\frac{1}{\tau_\nu} (f_i(\mathbf{x}, t) - f_i^{\text{eq}}(\mathbf{x}, t))$$

which achieves a relaxation of local densities towards an absolute equilibrium (at the macroscopic level). The time-scale τ_ν (in lattice unit) is related to the kinematic viscosity of the fluid by

$$\nu = \left(\tau_\nu - \frac{1}{2} \right) c_s^2 \Delta t$$

This simplification of the collision kernel is known as the BGK approximation in the kinetic theory of gas. The equilibrium function is given by

$$f_i^{\text{eq}}(\mathbf{x}, t) = w_i \rho(\mathbf{x}, t) \left(1 + \frac{\mathbf{u}(\mathbf{x}, t) \cdot \mathbf{c}_i}{c_s^2} + \frac{u_\alpha(\mathbf{x}, t) u_\beta(\mathbf{x}, t) (c_{i\alpha} c_{i\beta} - c_s^2 \delta_{\alpha\beta})}{2c_s^4} \right)$$

with the weight factors $w_0 = 4/9$, $w_{1\dots 4} = 1/9$ and $w_{5\dots 8} = 1/36$ for the D2Q9 lattice.

This discrete Lattice Boltzmann scheme is second-order accurate in Δx and compliant to the weakly-compressible Navier-Stokes equations with a third-order error in $\text{Ma} = |\mathbf{u}|/c_s$ as the lattice spacing vanishes, i.e. $\Delta x \rightarrow 0$.

As mentioned before, the pressure is directly accessible from the mass density: $p = \rho c_s^2$. The viscous stress is also obtained easily from the densities of particles by

$$\tau_{\alpha\beta}^{\text{visc.}} = -\frac{\nu}{\tau_\nu c_s^2 \Delta t} \sum_i c_{i\alpha} c_{i\beta} (f_i - f_i^{\text{eq}})$$

so that the total stress expresses as

$$\tau_{\alpha\beta} = -c_s^2 \sum_i f_i \delta_{\alpha\beta} - \frac{\nu}{\tau_\nu c_s^2 \Delta t} \sum_i c_{i\alpha} c_{i\beta} (f_i - f_i^{\text{eq}}) \quad (\text{A } 1)$$

Finally, let us mention that in the present context of turbulent flows, the single-relaxation-time BGK collision has been replaced by a multi-relaxation-time procedure based on central moments with an improved stability (Rosis 2016).

REFERENCES

- ALDOUS, D. & VAZIRANI, U. 1994 "go with the winners" algorithms. In *Proceedings 35th Annual Symposium on Foundations of Computer Science*, pp. 492–501.
- ARISTOFF, DAVID, LELIÈVRE, TONY, MAYNE, CHRISTOPHER G. & TEO, IVAN 2015 Adaptive Multilevel Splitting in Molecular Dynamics Simulations. *ESAIM: Proceedings and Surveys* **48**, 215–225.
- BEC, JEREMIE & KHANIN, KONSTANTIN 2007 Burgers Turbulence. *Physics Reports* **447** (1-2), 1–66, arXiv: 0704.1611.
- BOLHUIS, PETER G 2005 Kinetic pathways of β -hairpin (un) folding in explicit solvent. *Biophysical journal* **88** (1), 50–61.
- BOUCHET, FREDDY, LAURIE, JASON & ZABORONSKI, OLEG 2014 Langevin dynamics, large deviations and instantons for the quasi-geostrophic model and two-dimensional euler equations. *Journal of Statistical Physics* **156** (6), 1066–1092.
- BOUCHET, FREDDY, ROLLAND, JORAN & SIMONNET, ERIC 2019 Rare event algorithm links transitions in turbulent flows with activated nucleations. *Physical review letters* **122** (7), 074502.
- BREWER, TOBIAS, CLARK, STEPHEN R, BRADFORD, RUSSELL & JACK, ROBERT L 2018 Efficient characterisation of large deviations using population dynamics. *Journal of Statistical Mechanics: Theory and Experiment* **2018** (5), 053204.
- CÉROU, FRÉDÉRIC & GUYADER, ARNAUD 2007 Adaptive Multilevel Splitting for Rare Event Analysis. *Stochastic Analysis and Applications* **25** (2), 417–443.
- CÉROU, FRÉDÉRIC, GUYADER, ARNAUD & ROUSSET, MATHIAS 2019 Adaptive multilevel splitting: Historical perspective and recent results. *Chaos: An Interdisciplinary Journal of Nonlinear Science* **29** (4), 043108.
- CHERNYKH, A. I. & STEPANOV, M. G. 2001 Large negative velocity gradients in Burgers turbulence. *Physical Review E* **64** (2), 026306.
- DEL MORAL, PIERRE 2013 *Mean field simulation for Monte Carlo integration, Monographs on Statistics and Applied Probability*, vol. 126. CRC Press, Boca Raton, FL.
- DEMATTÉIS, GIOVANNI, GRAFKE, TOBIAS, ONORATO, MIGUEL & VANDEN-EIJNDEN, ERIC 2019 Experimental evidence of hydrodynamic instantons: The universal route to rogue waves. *arXiv preprint arXiv:1907.01320*.
- DEMATTÉIS, GIOVANNI, GRAFKE, TOBIAS & VANDEN-EIJNDEN, ERIC 2018 Rogue waves and large deviations in deep sea. *Proceedings of the National Academy of Sciences* **115** (5), 855–860.
- DONZIS, D. A. & SREENIVASAN, K. R. 2010 Short-term forecasts and scaling of intense events in turbulence. *Journal of Fluid Mechanics* **647**, 13–26.
- DÜBEN, P., HOMEIER, D., JANSEN, K., MESTERHAZY, D., MÜNSTER, G. & URBACH, C.

- 2008 Monte Carlo simulations of the randomly forced Burgers equation. *EPL (Europhysics Letters)* **84** (4), 40002.
- VAN ERP, TITUS S. & BOLHUIS, PETER G. 2005 Elaborating transition interface sampling methods. *Journal of Computational Physics* **205** (1), 157–181.
- ESCOBEDO, FERNANDO A., BORRERO, ERNESTO E. & ARAQUE, JUAN C. 2009 Transition path sampling and forward flux sampling. Applications to biological systems. *Journal of Physics: Condensed Matter* **21** (33), 333101.
- GARRAHAN, JUAN P., JACK, ROBERT L., LECOMTE, VIVIEN, PITARD, ESTELLE, VAN DUIJVENDIJK, KRISTINA & VAN WIJLAND, FRÉDÉRIC 2007 Dynamical first-order phase transition in kinetically constrained models of glasses. *Physical review letters* **98** (19), 195702.
- GIARDINÀ, CRISTIAN, KURCHAN, JORGE, LECOMTE, VIVIEN & TAILLEUR, JULIEN 2011 Simulating Rare Events in Dynamical Processes. *Journal of Statistical Physics* **145** (4), 787–811.
- GIARDINÀ, CRISTIAN, KURCHAN, JORGE & PELITI, LUCA 2006 Direct Evaluation of Large-Deviation Functions. *Physical Review Letters* **96** (12), 120603.
- GLASSERMAN, PAUL, HEIDELBERGER, PHILIP, SHAHABUDDIN, PERWEZ & ZAJIC, TIM 1998 A Look At Multilevel Splitting. In *Monte Carlo and Quasi-Monte Carlo Methods 1996*, pp. 98–108. Springer, New York, NY.
- GLASSERMAN, PAUL, HEIDELBERGER, PHILIP, SHAHABUDDIN, PERWEZ & ZAJIC, TIM 1999 Multilevel Splitting for Estimating Rare Event Probabilities. *Operations Research* **47** (4), 585–600.
- GRAFKE, TOBIAS, GRAUER, RAINER & SCHÄFER, TOBIAS 2013 Instanton filtering for the stochastic Burgers equation. *Journal of Physics A: Mathematical and Theoretical* **46** (6), 062002.
- GRAFKE, TOBIAS, GRAUER, RAINER & SCHÄFER, TOBIAS 2015 The instanton method and its numerical implementation in fluid mechanics. *Journal of Physics A: Mathematical and Theoretical* **48** (33), 333001.
- GRASSBERGER, PETER 2002 Go with the winners: A general monte carlo strategy. *Computer Physics Communications* **147** (1-2), 64–70.
- GRASSBERGER, PETER & OTHERS 1998 Perm: a monte carlo strategy for simulating polymers and other things. *arXiv preprint cond-mat/9806321*.
- GRIGORIO, L. S., BOUCHET, F., PEREIRA, R. M. & CHEVILLARD, L. 2017 Instantons in a Lagrangian model of turbulence. *Journal of Physics A: Mathematical and Theoretical* **50** (5), 055501.
- GURARIE, VICTOR & MIGDAL, ALEXANDER 1996 Instantons in the Burgers equation. *Physical Review E* **54** (5), 4908–4914.
- HIDALGO, ESTEBAN GUEVARA 2018 Breakdown of the finite-time and -population scalings of the large deviation function in the large-size limit of a contact process. *Journal of Statistical Mechanics: Theory and Experiment* **2018** (8), 083211.
- HUBER, G. A. & KIM, S. 1996 Weighted-ensemble Brownian dynamics simulations for protein association reactions. *Biophysical Journal* **70** (1), 97–110.
- KAHN, H. & HARRIS, T. E. 1951 Estimation of Particle Transmission by Random Sampling. *National Bureau of Standards* **12**, 27–30.
- KOSZTIN, IOAN, FABER, BYRON & SCHULTEN, KLAUS 1996 Introduction to the diffusion monte carlo method. *American Journal of Physics* **64** (5), 633–644.
- KRAICHNAN, ROBERT H. & CHEN, SHIYI 1989 Is there a statistical mechanics of turbulence? *Physica D: Nonlinear Phenomena* **37** (1), 160 – 172.
- LAFFARGUE, TANGUY, LAM, KHANH-DANG NGUYEN THU, KURCHAN, JORGE & TAILLEUR, JULIEN 2013 Large deviations of lyapunov exponents. *Journal of Physics A: Mathematical and Theoretical* **46** (25), 254002.
- LATT, JONAS, CHOPARD, BASTIEN, MALASPINAS, ORESTIS, DEVILLE, MICHEL & MICHLER, ANDREAS 2008 Straight velocity boundaries in the lattice boltzmann method. *Phys. Rev. E* **77**, 056703.
- LAURIE, JASON & BOUCHET, FREDDY 2015 Computation of rare transitions in the barotropic quasi-geostrophic equations. *New Journal of Physics* **17** (1), 015009.
- LECOMTE, VIVIEN & TAILLEUR, JULIEN 2007 A numerical approach to large deviations in

- continuous time. *Journal of Statistical Mechanics: Theory and Experiment* **2007** (03), P03004.
- LESIEUR, M. 2011 *Turbulence in Fluids : Stochastic and Numerical Modelling*. Springer.
- LESTANG, THIBAULT 2018 Numerical simulation and rare events algorithms for the study of extreme fluctuations of the drag force acting on an obstacle immersed in a turbulent flow. Theses, Université de Lyon.
- LESTANG, THIBAULT, BOUCHET, FREDDY & EMMANUEL, LÉVÉQUE 2019 A statistical mechanics approach to the study of extreme fluctuations of the time-averaged drag acting on an obstacle immersed in a turbulent flow In preparation.
- LESTANG, THIBAULT, RAGONE, FRANCESCO, BRÉHIER, CHARLES-EDOUARD, HERBERT, CORENTIN & BOUCHET, FREDDY 2018 Computing return times or return periods with rare event algorithms. *Journal of Statistical Mechanics: Theory and Experiment* **2018** (4), 043213.
- LOUVIN, HENRI, DUMONTEIL, ERIC & LELIÈVRE, TONY 2017 Three-dimensional neutron streaming calculations using adaptive multilevel splitting.
- MESTERHÁZY, DAVID, BIFERALE, LUCA, JANSEN, KARL & TRIPICCIONE, RAFFAELE 2013 Lattice monte carlo methods for systems far from equilibrium. *arXiv preprint arXiv:1311.4386* .
- MESTERHÁZY, DAVID & JANSEN, KARL 2011 Anomalous scaling in the random-force-driven burgers' equation: a monte carlo study. *New Journal of Physics* **13** (10), 103028.
- MORAL, PIERRE DEL 2004a *Feynman-Kac Formulae: Genealogical and Interacting Particle Systems with Applications*. New York: Springer-Verlag.
- MORAL, PIERRE DEL 2004b *Feynman-Kac Formulae: Genealogical and Interacting Particle Systems with Applications*. Springer-Verlag, New York.
- NEMOTO, TAKAHIRO, GUEVARA HIDALGO, ESTEBAN & LECOMTE, VIVIEN 2017a Finite-time and finite-size scalings in the evaluation of large-deviation functions: Analytical study using a birth-death process. *Phys. Rev. E* **95**, 012102.
- NEMOTO, TAKAHIRO, JACK, ROBERT L. & LECOMTE, VIVIEN 2017b Finite-size scaling of a first-order dynamical phase transition: Adaptive population dynamics and an effective model. *Phys. Rev. Lett.* **118**, 115702.
- RAGONE, FRANCESCO, WOUTERS, JEROEN & BOUCHET, FREDDY 2018 Computation of extreme heat waves in climate models using a large deviation algorithm. *Proceedings of the National Academy of Sciences* **115** (1), 24–29.
- ROLLAND, JORAN 2018 Extremely rare collapse and build-up of turbulence in stochastic models of transitional wall flows. *Physical Review E* **97** (2), 023109.
- ROLLAND, JORAN, BOUCHET, FREDDY & SIMONNET, ERIC 2016 Computing transition rates for the 1-d stochastic ginzburg–landau–allen–cahn equation for finite-amplitude noise with a rare event algorithm. *Journal of Statistical Physics* **162** (2), 277–311.
- ROLLAND, JORAN & SIMONNET, ERIC 2015 Statistical behavior of adaptive multilevel splitting algorithms in simple models. *Journal of Computational Physics* **283**, 541–558, arXiv: 1412.3362.
- ROSIS, ALESSANDRO DE 2016 Non-orthogonal central moments relaxing to a discrete equilibrium: A d2q9 lattice boltzmann model. *EPL (Europhysics Letters)* **116** (4), 44003.
- SUKOP, MICHAEL C. & JR, DANIEL T. THORNE 2006 *Lattice Boltzmann Modeling: An Introduction for Geoscientists And Engineers*. Springer-Verlag.
- TAILLEUR, JULIEN & KURCHAN, JORGE 2007 Probing rare physical trajectories with Lyapunov weighted dynamics. *Nature Physics* **3** (3), 203–207.
- TEO, IVAN, MAYNE, CHRISTOPHER G., SCHULTE, KLAUS & LELIÈVRE, TONY 2016 Adaptive Multilevel Splitting Method for Molecular Dynamics Calculation of Benzamidine-Trypsin Dissociation Time. *Journal of Chemical Theory and Computation* **12** (6), 2983 – 2989.
- TOUCHETTE, HUGO 2009 The large deviation approach to statistical mechanics. *Physics Reports* **478** (1-3), 1–69, arXiv: 0804.0327.
- VILLENT-ALTAMIRANO, M. & VILLENT-ALTAMIRANO, J. 1994 RESTART: a straightforward method for fast simulation of rare events. In *Proceedings of Winter Simulation Conference*, pp. 282–289.
- WOUTERS, JEROEN & BOUCHET, FREDDY 2016 Rare event computation in deterministic chaotic

- systems using genealogical particle analysis. *Journal of Physics A: Mathematical and Theoretical* **49** (37), 374002.
- YEUNG, P. K., ZHAI, X. M. & SREENIVASAN, KATEPALLI R. 2015 Extreme events in computational turbulence. *Proceedings of the National Academy of Sciences* **112** (41), 12633–12638, arXiv: <https://www.pnas.org/content/112/41/12633.full.pdf>.
- ZUCKERMAN, DANIEL M & CHONG, LILLIAN T 2017 Weighted ensemble simulation: review of methodology, applications, and software. *Annual review of biophysics* **46**, 43–57.