

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

Thibault Lestang,^{1, 2, *} Freddy Bouchet,^{1, †} and Emmanuel Lévéque^{2, ‡}

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

² *Univ 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*

Abstract

This study evaluates the relevance of importance sampling techniques to optimize the numerical simulation of extreme mechanical efforts exerted by a turbulent flow impinging onto a bluff body. The main idea behind importance sampling is here to replace a long simulation by a set of shorter simulations running in parallel with dynamics that are replicated or abandoned in order to sample large events more frequently. Such techniques have already proved their efficiency for simulating extreme events in... with significant run-time savings. Application to turbulent fluid-mechanical systems remains mainly open. 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 important sampling algorithms, namely Practical evidence is given that these algorithms allow for the correct sampling of rare large-amplitude drag events and the estimation of return times with a reduced computational effort... However, limitations... pitfalls... Importantly, such techniques also provide access to the flow scenario yielding extreme events. In the present case, it is found that... blocking effect

I. INTRODUCTION

// Recall the previous success of importance sampling techniques with references. // Explain why interesting to test application for turbulent systems. // To what extent a turbulent system is different from previous case studies? //

Turbulent flows are important in a variety of natural phenomena, industrial and civil applications with the characteristic feature to spontaneously develop intense and sporadic fluid motions associated with enormous internal forces. In the turbulence literature, “extreme” can refer to fluctuations that deviate from the mean value by $O(10)$ standard deviations. // true ? // The nature of such extreme dynamical events and their statistics are of crucial interest to predict excessive mechanical efforts that may cause damages and threaten the structural integrity of embedded structures.

* thibault.lestang@ens-lyon.fr

† freddy.bouchet@ens-lyon.fr

‡ emmanuel.leveque@ec-lyon.fr

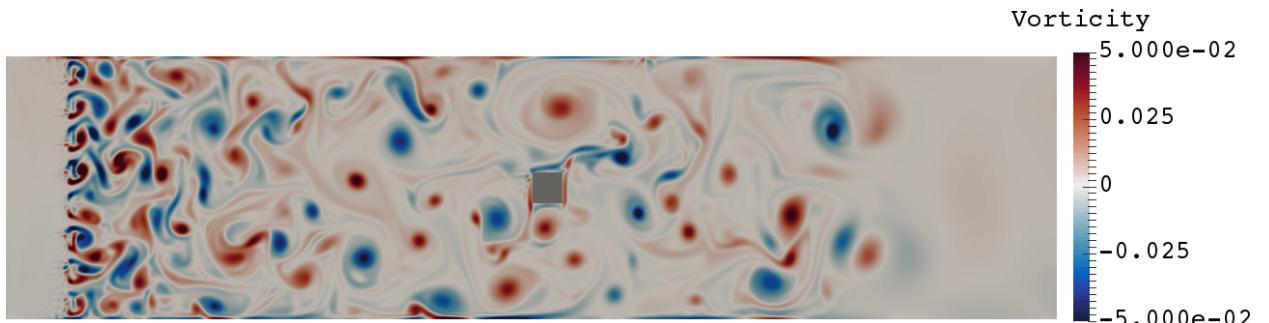


Figure 1: Our case study is a grid-generated turbulent flow impinging a squared obstacle located at the center of a channel (in two dimensions). The vorticity is displayed with an arbitrary colormap from blue (negative values) to red (positive values). The Reynolds number based on the grid spacing is $Re = \dots$ // no color bar, show the grid //

From the viewpoint of dynamical systems, turbulence in fluids has to do with non-linearity and strong departure from statistical equilibrium. The use of perturbative methods in identifying resonant interactions (among degrees of freedom) leading to extreme fluctuations is not successful. In this context, simulation offers a practical means to gain insight into the extreme events. However, their numerical sampling requires long simulation since these events are rare as well.

// Our motivations and expectations. //

II. DESCRIPTION OF THE CASE STUDY

The drag exerted by a grid-generated turbulent flow onto a squared obstacle is considered as our representative case study. The flow is contained in a long channel as illustrated in 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 make possible a systematic study. Nevertheless, this fluid-mechanical system exhibits the characteristic features of turbulent fluid-structure interaction. // true even if 2d? // In the present situation, the obstacle does not deform or move. Through this specific setting, our main motivation is to evaluate the operability of importance sampling techniques to capture extreme events with a significant run-time savings.

A constant parabolic profile is imposed for the streamwise velocity at the inlet of the channel and turbulence is generated by means of a grid. At the outlet of the channel, the flow is progressively damped within a *sponge layer* [? ?] where the viscosity is artificially enhanced.

Turbulent eddies are generated in the near-wake of the grid and carried downstream. // grid is usually characterized by the spacing and the mesh size, are they equal here? // They interact with each other and grow in size as expected in two-dimensional turbulence. The size of the eddies that eventually hit the square (at the center of the channel) is comparable to the size of this obstacle, which results in strong fluctuations of the drag acting on the square.

The flow dynamics are computed by the lattice Boltzmann (LB) method [? ?]. While standard numerical methods rely on the discretization of the Navier-Stokes equations, the LB method emerged in the late 80s as a conceptually different approach. Fluid motions are simulated at a mesoscopic level that is intermediate between the microscopic and the macroscopic. Capturing the kinetic behavior of collections of fluid particles distributed on a lattice is here preferred to solving non-linear PDEs. This seems crazy, however, most details at the mesoscopic level play actually no role at the macroscopic level. Therefore, kinetic equations simpler than the original Boltzmann equation may be designed by retaining only the basic features that pertain at the macroscopic level. This is, in short, the rationale behind the LB approach. In our situation, the LB method has been chosen principally for its outstanding computational efficiency. More precisely, the fluid is viewed as populations 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. 2). 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.

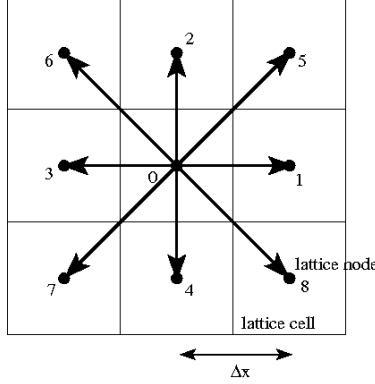


Figure 2: Sketch of the D2Q9 lattice. Particles move exactly from a lattice node towards one of its nine neighbours (including the node itself) during one time step. By definition, the lattice spacing is related to the time step by $\Delta x/\Delta t = \sqrt{3}c_s$, where c_s is interpreted as a speed of sound.

The complexity of the flow emerges from the repeated application of simple rules of streaming and collision. Therefore, the LB scheme 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 neighboring node, is consecutive to 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 can be viewed as the relaxation of local densities towards an absolute equilibrium at the macroscopic level; the time-scale τ_ν 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 (assuming Einstein summation convention) 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. Finally, let us mention that this discrete LB 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. $\Delta x \rightarrow 0$.

```
// algorithm for boundary conditions // computation of the stress with LBM // grid
resolution, number of grid points in each direction, Reynolds, Mach numbers, etc. // low
Mach number approaches the nearly-incompressible limit
```

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

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

A. 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 and reads

$$f_d(t) = \int_{\mathcal{S}} \boldsymbol{\tau}_{x\beta}(\mathbf{x}, t) d\mathcal{S}_\beta(\mathbf{x}) \quad (2)$$

where \mathcal{S} is the surface of the obstacle. In our situation, the viscous stress makes a negligible contribution to the drag. This latter therefore results mostly from pressure forces, which are closely related to the distribution of velocity gradients in the vicinity of the obstacle (in the nearly-incompressible limit). 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 may be viewed as the pressure difference

$$f_d(t) = p_{fb}(t) - p_b(t) \quad (3)$$

where $p_{fb}(t)$ and $p_b(t)$ denote the pressure integrated over the forebody and the base of the obstacle, respectively.

The typical timescale of variation of the drag therefore corresponds to the timescale of pressure fluctuations in the vicinity of the obstacle. An estimate of such timescale is the

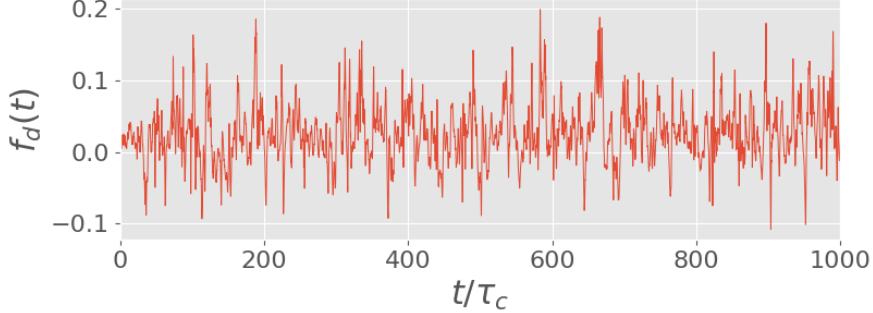


Figure 3: Typical temporal evolution of the drag acting on the square cylinder pictured in figure 1. The correlation time τ_c is defined further in this text and correspond to the typical fluctuation timescale of the drag.

turnover time

$$\tau_0 = \frac{R}{U}, \quad (4)$$

where R is the diameter of the cylinder and U the mean-flow velocity. Figure 4 illustrates the time evolution of the drag acting on the square obstacle over a few thousands turnover times. The evolution over a few turnover times of the vorticity field around the obstacle is displayed in figure ???. One can see that the interaction of the incoming flow with the square obstacle primarily results in boundary layers along the top and bottom boundaries. However, the vorticity corresponding to these boundary layers only have a small impact on the pressure field in the vicinity of the obstacle, because it is rapidly advected downstream by the mean-flow. As a consequence, the base pressure varies much less than the forebody pressure, and typical drag fluctuations mostly result from the fluctuating pressure at the front of the obstacle.

In the following section III, we highlight a very different phenomenology for *extreme* fluctuations of the drag acting on the obstacle. In particular, we show that extreme fluctuations mainly result from fluctuations of the base pressure, whereas forebody pressure fluctuations play a smaller role.

B. The drag random process: Probability Density Function, time correlations

In the following we consider the drag force acting on the obstacle as a scalar random process $f_d(t)$. In order to estimate its statistical properties, we performed a simulation of

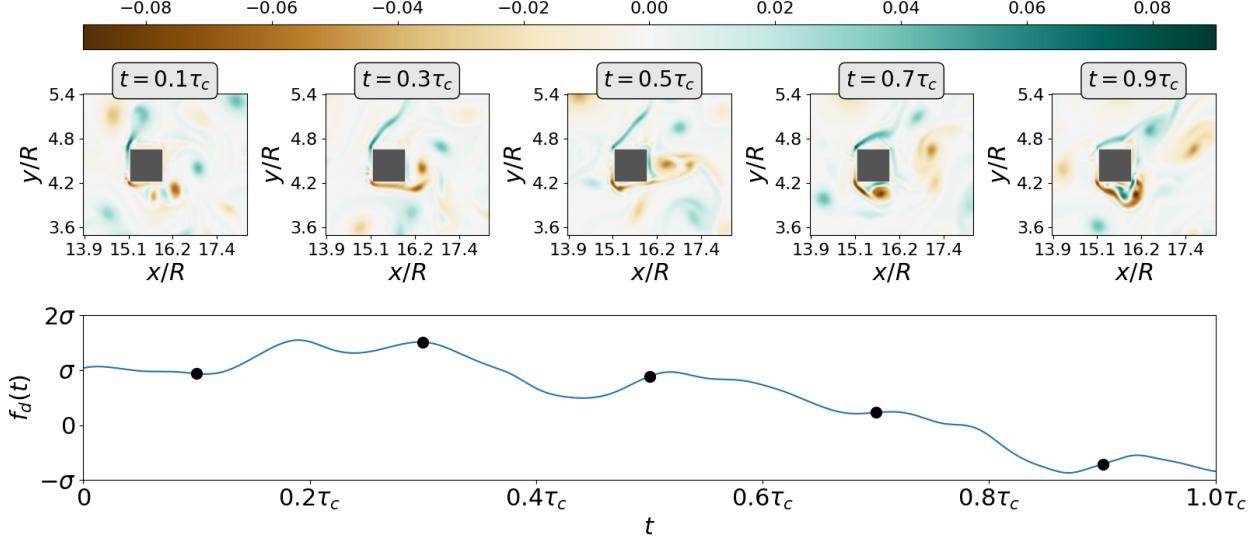


Figure 4: Vorticity dynamics corresponding to typical drag fluctuations over one correlation time unit τ_c . Negative vorticity is generated along the bottom boundary, and advected away from the the obstacle by the mean flow.

the dynamics over a duration $T_{tot} = 4 \times 10^6 \tau_0$. In the following of this paper we refer to this long simulation as the *control run*. Furthermore, we denote by \bar{f}_d and σ the average and variance computed over the control run, respectively. Figure 5 displays an estimate of the PDF for the drag process, computed on the basis of the control timeseries $\{f_d(t)\}_{0 \leq t \leq T_{tot}}$ obtained from the control run. One can see that the tails of this distribution, that describe the probability of very rare events, can be very well modelled by an exponential PDF of the form $p(f_d) = \lambda^{-1} e^{-\lambda f_d}$. In addition, figure 5 displays an estimate of the PDF of the process describing the drag acting on a control surface equivalent to the surface of the obstacle, in a flow *without obstacle*. This different setup is illustrated in figure ???. In this case, the corresponding PDF estimate for the drag process is quasi-symmetric and, more importantly, does not display exponential tails. We conclude that the interation of the flow with the square obstacle is responsible for the exponential tails of the PDF of the drag process. In section III C, we relate this very particular shape of the PDF describing extreme drag fluctuations to the phenomenology of extreme fluctuations of the time-average of the drag process.

Lastly, figure ?? illustrates that the drag process has exponentially decaying time correlations. In the following we call *correlation time*, denoted by τ_c , the typical time it takes

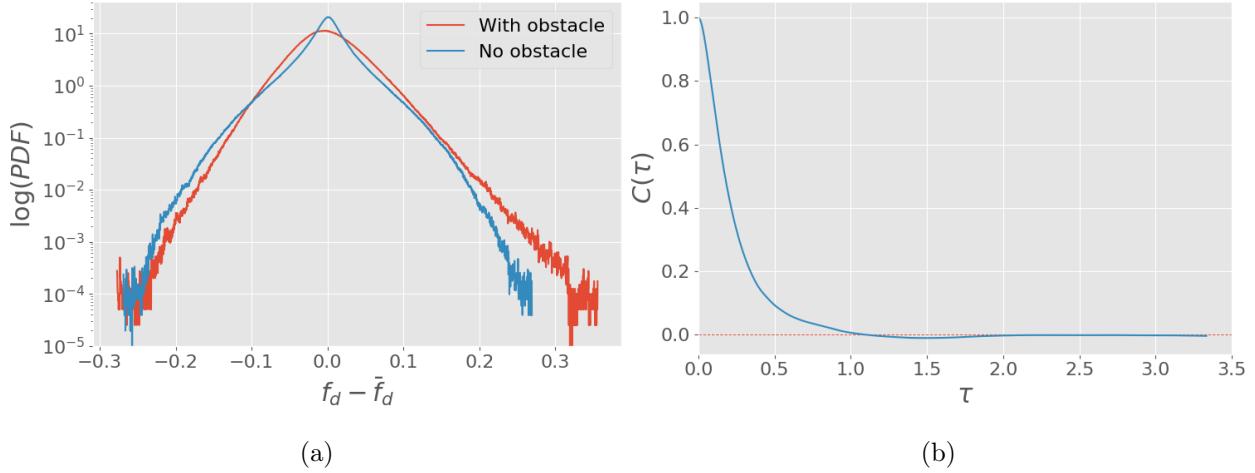


Figure 5: **(a)** PDF describing the statistics of the instantaneous drag fluctuations, computed on the basis of the control timeseries. **(b)** Autocorrelation function for the instantaneous drag f_d , defined as $(\mathbb{E}[f_d(t + \tau)f_d(t)] - \mathbb{E}[f_d]^2)/\sigma^2$, computed over the control run. It shows the exponential decay of the correlations over time.

for the drag process to decorrelate. We observe that $\tau_c \approx \tau_0$. This means that the drag decorrelates after only a few number of eddies interacted with the square obstacle.

III. DIRECT SAMPLING OF EXTREME FLUCTUATIONS OF THE DRAG ACTING ON A SQUARE CYLINDER IN TWO-DIMENSIONAL TURBULENCE

We now turn to a description of the phenomenology of extreme fluctuations of the drag that acts on the square cylinder mounted in the flow described in section II.

How rare is a fluctuation $f_d \geq a$ can be determined by computing its *return time* $r(a)$. It is defined as the average time between two consecutive occurrences of a fluctuation $f_d \geq a$. Extreme drag fluctuations are *rare events*, that is fluctuations for which $r(a) \gg \tau_c$. In this context, the occurrence of the fluctuations follows a Poisson process and [?]

$$r(a) \underset{a \rightarrow \infty}{\propto} \frac{1}{\mathbb{P}(f_d \geq a)} \propto e^{\alpha a} \quad (5)$$

where α is the rate describing the positive tail of the drag PDF, see figure 5. Figure 6 illustrates the fluctuation amplitude a as a function of the return time $r(a)$, for the drag process $f_d(t)$. Such plot can be obtained from a sample timeseries following a simple procedure [?].

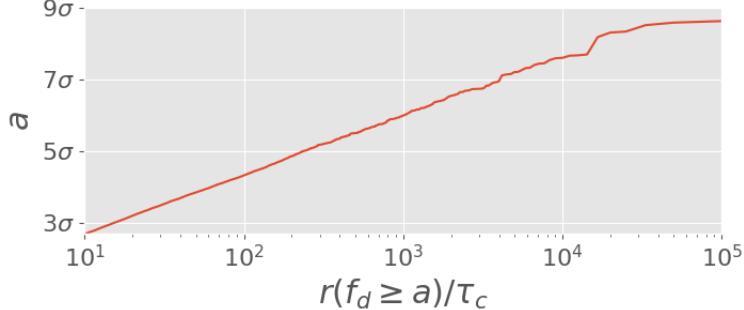


Figure 6: Drag fluctuation amplitude as a function the corresponding return time. Here f_d denotes the centered and reduced drag with $f_d = 0$ and $\sigma = 1$.

In the following we discuss extreme fluctuations of both the instantaneous drag f_d and time-averaged drag with return times larger than $10^4\tau_c$. In this section extreme drag fluctuations are simply sampled by simulating the flow over a very long duration. This approach is referred to as *direct sampling*, as opposed to approaches involving the use of rare events algorithms, discussed in section IV. Such approach is made possible by the relative simplicity of the flow described in section II. The control run mentioned in the previous section, spanning $T_{tot} = 10^6\tau_c$, could be achieved in a wall-clock time of a few weeks and gives access to extreme fluctuations $f_d \geq a$ with a return time $r(a) \leq 10^6\tau_c$.

A. Extracting extreme drag fluctuations from a very long timeseries

On the basis of the control timeseries $\{f_d(t)\}_{0 \leq t \leq T_{tot}}$, we tracked every drag fluctuation with an amplitude greater than a fixed threshold amplitude a . The value of this threshold was chosen so that the return time of fluctuations with an amplitude larger than a is $T_{tot}/100$. Figure ?? displays the fluctuation amplitude a as a function of the return time. According to Figure ??, the threshold amplitude a is set to 7.6σ with σ the standard deviation of the drag process. We denote by (t^*, f_d^*) a fluctuation of peak amplitude f_d^* occurring at time t^* , as illustrated in figure ???. From the control timeseries, we identified 104 independent extreme drag fluctuations whith an amplitude greater than 7.6σ , *i.e.* with a return time greater than $10^4\tau_c$. In the following sections III B 1 and III B, we describe the phenomenology of extreme drag fluctuations on the basis of this set of events.

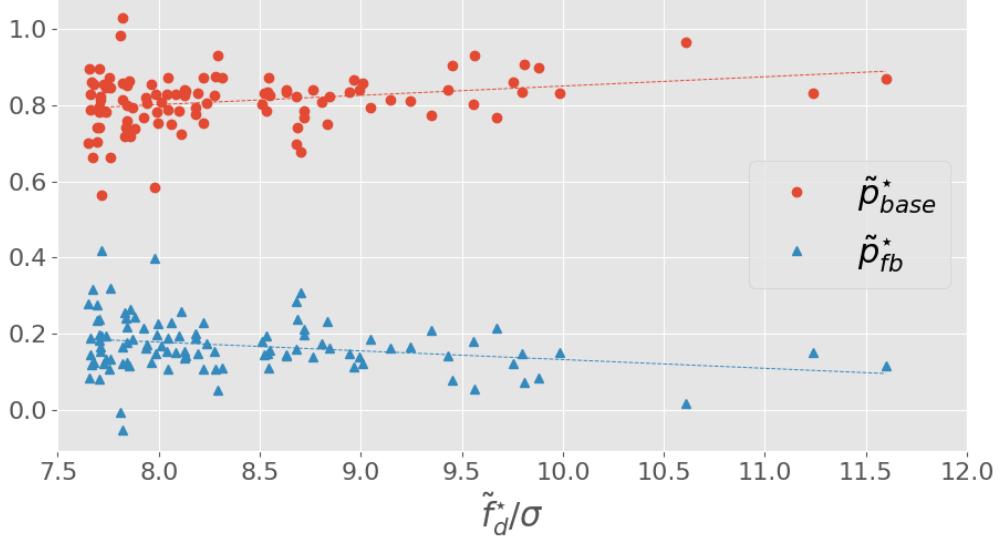


Figure 7: Contribution of the forebody and base pressure fluctuation to the overall drag fluctuation for the 104 sampled extreme drag fluctuations in the control run. To each of these fluctuation correspond a unique value of the drag f_d^* , and a unique pair $(\tilde{p}_{base}^*, \tilde{p}_{fb}^*)$. The dashed lines are first order least square fits of the data. They highlight that, as the value of the fluctuation increases, so is the relative contribution of the forebody pressure fluctuation.

B. Instantaneous drag

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

In section II, we pointed out that typical drag fluctuations mainly originate from fluctuations of the forebody pressure, *i.e.* pressure fluctuations resulting from the upstream turbulent flow. We now investigate the relative contribution of forebody and base pressure fluctuations to the overall drag fluctuations, in the case of the sampled *extreme* drag fluctuations. Let (t^*, f_d^*) be an extreme drag event. In the following we consider the effective fluctuation $\tilde{f}_d^* = f_d^* - \bar{f}_d$. It can be further decomposed into

$$\tilde{f}_d^* = \tilde{p}_{fb}^* - \tilde{p}_{base}^*, \quad (6)$$

where \tilde{p}_{fb}^* and \tilde{p}_{base}^* denote the fluctuation of the forebody pressure and base pressure, respectively, that is $\tilde{p}_{base}^* = p_{base}^* - \bar{p}_{base}$. Figure ?? shows the relative contributions $\tilde{p}_{base}^*/\tilde{f}_d^*$ and $\tilde{p}_{fb}^*/\tilde{f}_d^*$ to the overall drag fluctuation for the base and forebody pressure fluctuations,

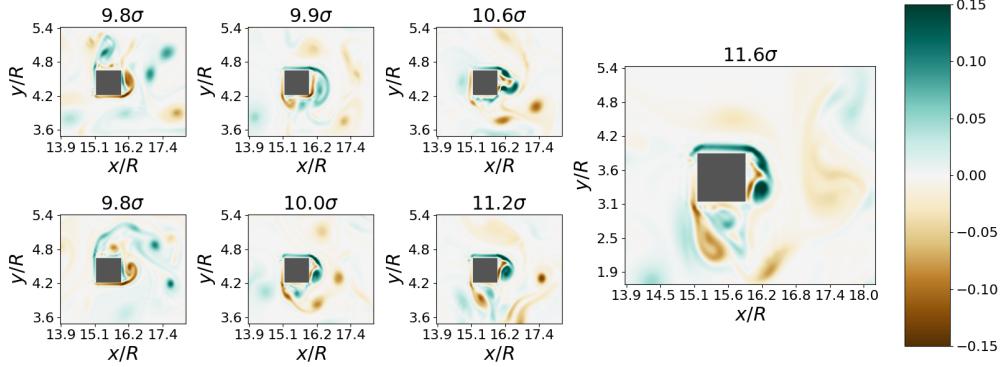


Figure 8: Vorticity field at $t = t^*$ for some of the highest drag fluctuations in the control timeseries. The extreme value for the drag results from the concentration of vorticity in the vicinity of the base of the obstacle, causing a pressure drop.

respectively.

It shows that among the 104 sampled extreme events, the base pressure fluctuation typically contributes for 80% of the overall drag fluctuation. By contrast with typical drag fluctuations, extreme fluctuations of the drag acting on the square cylinder are dominated by the variation of the pressure in the vicinity of the base of the obstacle. In addition, figure ?? hints that, the larger the fluctuation, the more important is the contribution of the base pressure fluctuation, relatively to the forebody pressure fluctuation.

2. Dynamical aspects of extreme drag fluctuations

We now turn to a description of the flow dynamics that correspond to extremely high values of the drag acting on the square obstacle.

Figure ?? displays the average of the drag timeseries corresponding to the 104 extreme events sampled in the control run, in the vicinity of the fluctuation. It illustrates that extreme drag fluctuations are very local in time, with a lifetime of roughly one correlation time unit. Starting from typical values, extreme drags are typically reached in less than a correlation time. Moreover, such very high drag levels do not persist over time. To understand why it is so, it is useful to visualize the corresponding flow fields.

To begin with, figure ?? displays the vorticity field corresponding to the peak drag value, for a subset of the extreme events sampled in the control simulation. We observe that, in each case, extreme drag fluctuations correspond to very similar flow configurations. More

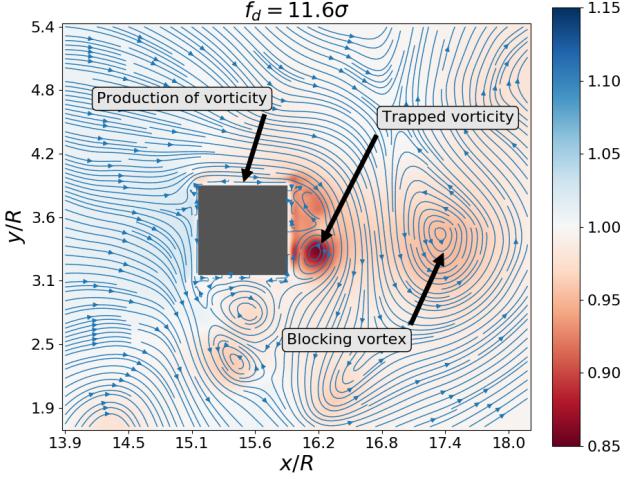


Figure 9: Pressure field and velocity streamlines in the vicinity of the obstacle at $t = t^*$. Because of the blocking vortex, vorticity produced along the top boundary of the obstacle has no other choice but to accumulate at the base of the obstacle.

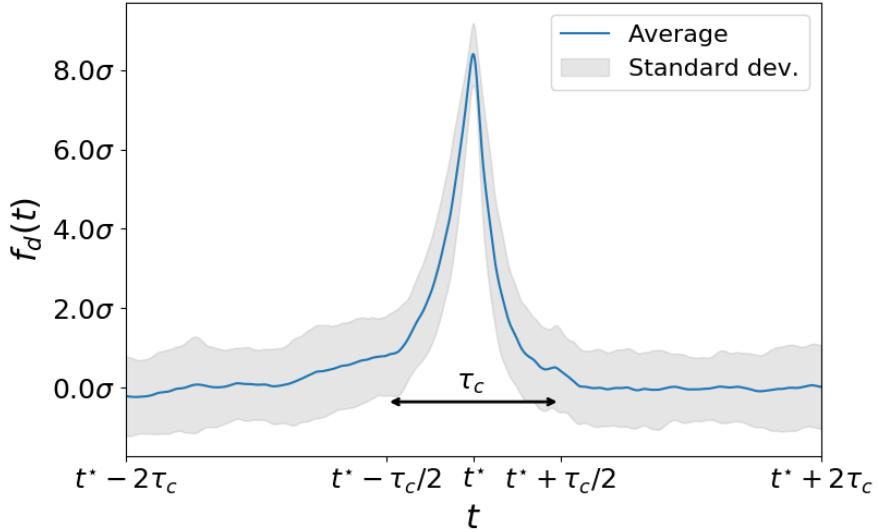


Figure 10: Average drag timeseries in the vicinity of an extreme drag fluctuation at $t = t^*$. The blue continuous line results from an average over the 104 extreme events sampled in the control run. This illustrates that extreme drag variations are very local in time, with a typical lifetime of one correlation time unit τ_c .

precisely, a region of strong vorticity is located in the vicinity of the base of the obstacle, where vorticity levels typically reach twice as much as the typical vorticity fluctuations displayed in figure ???. This vorticity results from a boundary layer, along either the top or bottom side of the obstacle. Such extreme vorticity is responsible for a significant pressure

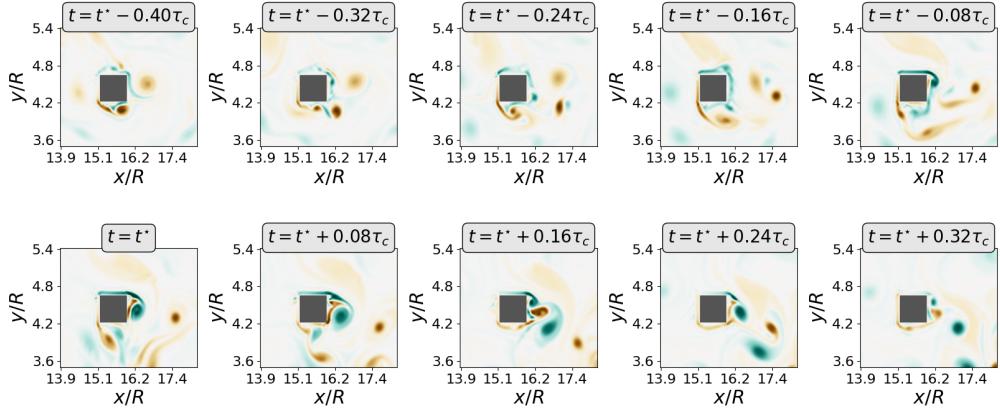


Figure 11: Vorticity dynamics associated to an extreme fluctuation of the drag acting on the square obstacle at $t = t^*$.

drop in the vicinity of the base of the obstacle, which in turn leads to an extreme value of the drag, see figure ??.

How are such vorticity reached ? In contrast with the typical flow dynamics, illustrated in figure ??, the vorticity generated within the boundary layers is not directly advected by the mean flow. Figure ?? shows that, instead, this vorticity is constrained in the vicinity of the base of the obstacle. Such constraint results from a large vortical region located roughly one obstacle length of the base, that prevents vorticity to be advected downstream.

Figure ?? illustrate the time evolution of the vorticity field, in the vicinity of the drag fluctuation, for the event depicted in figure ???. From $t^* - 0.4\tau_c$ to $t^* - 0.16\tau_c$, a strong negative eddy is generated along the bottom boundary layer of the obstacle and advected downstream by the mean flow. This leads to a region of negative vorticity at roughly one obstacle length of the base of the obstacle, constraining the positive vorticity generated along the top boundary. However, it can be seen in figure ?? that the positive vorticity only accumulates for a short duration. The blocking negative vortex is in turn advected downstream, thus freeing the trapped positive vorticity.

The lifetime of an extreme drag fluctuation therefore corresponds to the duration over which the blocking vortex remains in the vicinity of the base. Consistently with figure ??, this duration is roughly one correlation time unit τ_c .

We found that 80% of the extreme events sampled from the control timeseries can be related to very similar dynamics. In order to identify such events, we monitored the averaged

shear along the top or bottom boundary

$$\bar{\gamma} = \frac{1}{L} \int_{S_b} \frac{\partial u_x(\mathbf{x})}{\partial y} d\mathbf{x}, \quad (7)$$

where L denotes the diameter of the cylinder, u_x the longitudinal component of the velocity field and S_b the surface of either the top or the bottom boundary. Figure 12 shows $\bar{\gamma}$ as a function of the instantaneous drag f_d , for $t^* - 2\tau_c \leq t \leq t^* + 2\tau_c$, for each of the sampled events that follow the previously described dynamics.. For $t^* - 2\tau_c \leq t \leq t^* - \tau_c$ and $t^* + \tau_c \leq t \leq t^* + 2\tau_c$, paths concentrate in the region describing typical values for both $\bar{\gamma}$ and f_d . As illustrated in figure ??, the drag abruptly varies for $t^* - \tau_c \leq t \leq t^* + \tau_c$. As a consequence, paths in the $(f_d, \bar{\gamma})$ space display excursions to atypical values for both $\bar{\gamma}$ and f_d . 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.

The remaining 20% of the sampled events correspond to slightly different dynamics. In this case the vorticity responsible for the base pressure drop is not created along the top or bottom boundary, but directly through viscous shear alongside the base boundary of the obstacle. This viscous shear is induced by a large vortex detached from the obstacle. For a more detailed description of such dynamics, see the chapter 3 of ??.

C. Extreme fluctuations of the time-averaged drag

In section III B, we discussed the phenomenology of rare events corresponding to extremely high values of the drag acting on the square obstacle mounted in the flow described in section II. In particular, it was pointed out that such extreme drag fluctuations have a lifetime of, roughly, one correlation time unit τ_c . However, In many applications, this duration is much smaller than the timescale of interest. Consider for instance the interaction of a deformable structure with a turbulent flow: the typical response time may be much larger than the lifetime of drag fluctuations.

In such cases, a relevant observable is the *time-averaged* drag

$$F_T(t) = \frac{1}{T} \int_t^{t+T} f_d(t) dt, \quad (8)$$

where f_d denotes the instantaneous drag and $T \leq \tau_c$ a timescale of interest. In the following we consider the case where $T \gg \tau_c$, typically $T = 10\tau_c$. In this context, a fluctuation of the

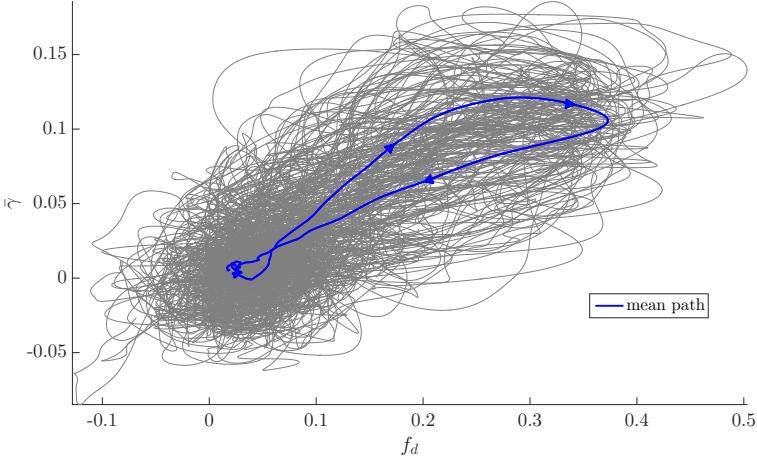


Figure 12: Evolution of the shear alongised the top or bottom boundary of the obsatcle as a function of the drag, for $t^* - 2\tau_c \leq t \leq t^* + 2\tau_c$. To each oneof the grey lines correspond a unique extreme event. The blue line is the average over the 104 extreme event sampled in the control run.

average $F_T(t)$ is related to roughly T/τ_c independent fluctuations of the *instantaneous* drag f_d , over the time interval $[t; t + T]$.

What is the phenomenology leading to extreme values of $F_T(t)$? Do an exceptionally large value of the average drag result from a single exceptionally large value of the instantaneous drag (case (1)) ? Or instead from a succession of rather typical fluctuations, however all in the same direction (case (2)) ? From the control timeseries, see section II B, one can compute a control timeseries $\{F_T(t)\}_{0 \leq t \leq T_{tot}-T}$. It is then possible to identify extreme positive fluctuations for which the time-averaged F_T drag exceeds a fixed threshold a , in the very same way as described in section III A. Setting $a = 5.2\sigma_T$ leads to 84 independent events, with σ_T the standard deviation of the random process describing the time evolution of F_T .

Figure ?? displays the timeseries $\{f_d(t^*)\}_{t^* \leq t \leq t^*+T}$ for four of the sampled extreme fluctuations sampled in the control timeseries for F_T . It illustrates that the phenomenology of the extreme fluctuations of the time-averaged drag can neither be reduced to case (1) nor case (2). Indeed, both cases are featured in figure ??, along with intermediate cases where the very large value of the drag results from both a very large fluctuation and a large number of typical, however all positive, fluctuations of the instantaneous drag f_d .

This marginal phenomenology can be connected to the exponential shape of the tail of the PDF describing extreme positive drag fluctuations [?]. An estimate of this PDF is displayed in figure 5. Let X be a random variable whose PDF is denoted \mathbb{P} and standard deviation σ_X . Considering an extreme positive value a of $S_N = \sum_{n=1}^N NX_n$, the probability p_1 (resp. p_2) of case (1) (resp. case (2)) writes:

$$p_1\left(\sum_1^N X_n = a\right) \approx \mathcal{P}\left(\frac{a}{N}\right)^N \quad \text{and} \quad p_2\left(\sum_1^N X_n = a\right) \approx \mathcal{P}(a) \quad (9)$$

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

$$\frac{p_2}{p_1} \underset{a \rightarrow \infty}{\sim} C \left(e^{-\alpha \frac{a}{N}}\right)^N e^{-\alpha a} = 1. \quad (10)$$

IV. RARE EVENT ALGORITHMS

In the limit of very rare events and/or very complex dynamics such as turbulent flows in industrial contexts, a direct sampling approach is unrealistic. Indeed, according to equation (??), the return time of a fluctuation $f_d \geq a$ verifies $r(a) \propto e^{\alpha a}$. As a result, the computational cost required to sample events $f_d \geq a$ scales like $e^{\alpha a}$.

In this section we discuss the application of **rare event algorithms** to the numerical sampling of extreme drag forces on immersed objects, on the example of the flow presented in section II. The purpose of rare events algorithms is to make it possible to sample rare events for a computational cost well inferior to their return time. In sections IV A and IV B we consider the Adaptive Multilevel Splitting and Giardina-Kurchan-Tailleur-Lecomte algorithms, respectively. Both algorithms rely on the simulation of an ensemble of N trajectories $\{\mathbf{x}_n(t)\}_{1 \leq n \leq N, 0 \leq t \leq T_a}$, where $\{\mathbf{x}(t)\}_{0 \leq t \leq T_a}$ is a formal notation for a trajectory of duration T_a in the phase space describing the dynamical system. In this case the dynamical system is the flow described in section II. Over the iterations of the algorithm, trajectories are duplicated or discarded from the ensemble according to precise selection rules that bias the sampling towards the rare events of interest. Such selection rules are designed in such a way that the sampling bias is known at each iteration of the algorithm. In other words, generated trajectories are given a statistical weight, which knowledge allows for the computation of both their probability and expectation values of any trajectory observables.

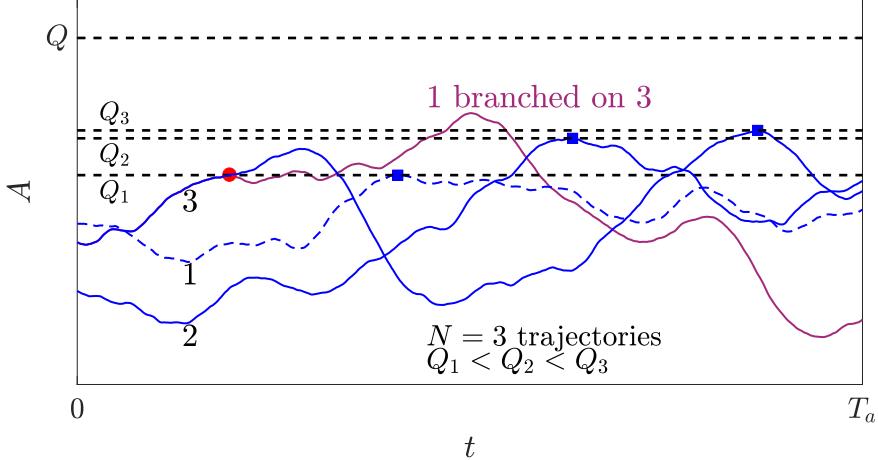


Figure 13: Illustration of one selection-mutation step in the AMS algorithm for the computation of the probability that an observable $A : \mathbb{R}^d \rightarrow \mathbb{R}$ reaches values larger than Q over a trajectory of duration T_a . The initial ensemble of blue trajectories is first generated. These trajectories are independant and can be simulated in parallel. On the basis of the respective maxima Q_1 , Q_2 and Q_3 , the trajectory with the lowest maximum is discarded from the ensemble (dashed blue line). Among the two remaining trajectories, trajectory 3 is chosen at random and copied until Q_1 . It is then freely simulated from the branching point to T_a . In case of deterministic dynamics, as small perturbation is introduced at the branching. This procedure is typically iterated J times, or until all trajectories go beyond a fixed threshold Q .

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

The Adaptive Multilevel Splitting [?] inherits from the basic ideas of *splitting* algorithms [?]: the sampling of a rare event is made easier by splitting the corresponding dynamical path into a sequence of events that can be sampled with a higher probability [? ?]. Even though a large body of variants have been proposed since the early 50s, the AMS appears as the most advanced formulation of multilevel splitting to date (**Sans doute un peu «too-much».** **Avec cette phrase je voudrais justifier que l'on s'intéresse à l'AMS et pas à un autre algo.**). In recent years, it was applied to several problems, for instance the computation of the dissociation time of biomolecules [?] or the simulation of rare relaminarizations in a stochastic model of wall turbulence [?].

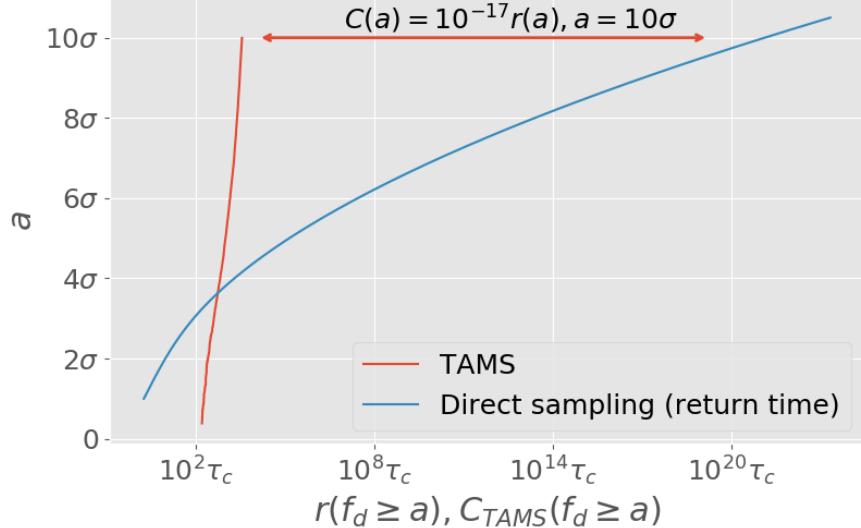


Figure 14: Illustration of the efficiency of the TAMS algorithm with respect to direct sampling. The orange line represents the evolution of the maximum over resampled trajectories as a function of the computational cost C_{TAMS} (iterations of the selection-mutation step). The solid blue line is the analytical solution for the return time of amplitude a for the OU process [?].

In the present paper we use a variant of the AMS algorithm called the Trajectory Adaptive Multilevel Splitting (TAMS). The TAMS relies on the definition of a time-dependant *score function* $\xi(\mathbf{x}(t), t)$, as well as an iterative procedure starting from an ensemble $\{\mathbf{x}_n(t)\}_{1 \leq n \leq N, 0 \leq t \leq T_a}$ of independant trajectories. Each iteration consists in discarding the trajectories with the lowest maxima of the score function ξ over the whole trajectory of duration $[0; T_a]$. Discarded trajectories are resampled based on member of the ensemble of trajectories that achieved a higher value of the score function at some point of their history. As an illustration, a particular iteration of the TAMS algorithm is sketched in figure ???. For further details about the algorithm and corresponding mathematical results, see [?] and references therein.

We first describe results of the TAMS on a simple case. We temporarily let fluid dynamics aside and consider a one-dimensional OU process

$$\dot{x} = -x + \sqrt{2\epsilon}\eta(t) \quad (11)$$

We apply the TAMS to simulate trajectories $\{x(t)\}_{0 \leq t \leq T_a}$ with $\max_{0 \leq t \leq T_a} x(t) \geq a$ with a very large with respect to typical fluctuations of $x(t)$. For the sake of simplicity, a single tra-

jectory is resampled at each iteration of the TAMS. We denote by a_j the maximum of the resampled trajectory at iteration j . The computational cost of the algorithm after iteration j is the computational cost of simulating the N initial trajectories, as well as the resampling of the j trajectories. Figure ?? compares the typical computational cost required to sample fluctuations above a given amplitude a using the TAMS and a direct sampling approach. In the latter, the typical computational cost is simply the return time $r(a)$. illustrates that the successive resamplings of the TAMS algorithm leads to trajectories displaying extreme fluctuations $x \geq a$. More importantly, these very rare trajectories are sampled for a computational cost that several order of magnitude lower than the return time of the corresponding fluctuations.

Undoubtedly, an OU process is an oversimplified dynamics to showcase the efficiency of the AMS and TAMS. Indeed, the state space is one-dimensional and the choice of the score function poses no questions: it is simply the observable x itself. In addition, the noise term in (11) has no correlations in time, which means that newly generated trajectories will quickly separate from their ancestors.

Can the TAMS achieve similar results when applied to complex chaotic dynamics, such as turbulent flows around objects ? To answer this question, we consider again the two-dimensional test flow introduced in section II. Our aim is to sample trajectories that display extreme fluctuations of the drag f_d acting on the square obstacle. However, by contrast with a simple OU process, the phase space is highly-dimensional and the dynamics complex. As a consequence the choice of the score function $xi(\mathbf{x}(t), t)$ is not so clear. In the following we use the TAMS with $\xi \equiv f_d$, which is the most simple choice for the score function. For such complex dynamics however, there is no reasons for the drag itself to be a good measure of how far is a given state to regions of phase space where the drag is extreme. Unfortunately, the choice of better score functions rely on a deep understanding of the physics of the rare events of interest, which is precisely what we are trying to achieve.

We apply the TAMS with $N = 256$ initial trajectories, with a duration $T_a = 20\tau_c$. Similarly to figure 14, figure ?? displays the maximum drag achieved by resampled trajectories. In addition, figure ?? shows the distribution of the maximum achieved drag for the N initial trajectories. Over the first iterations of the algorithm, trajectories with the lowest maximum of the score function are discarded, and new trajectories with higher maximum are resampled. These maxima are depicted by the stars in figure ???. However, one can see that

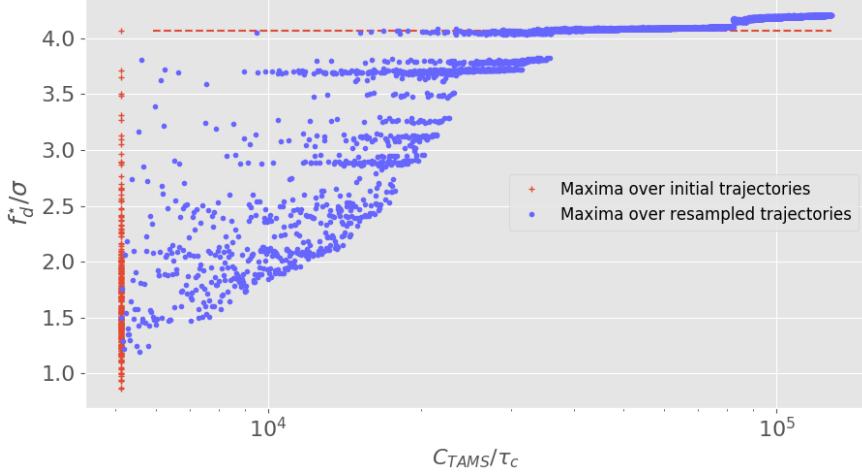


Figure 15: Maximum of the instantaneous drag along the resampled trajectories as a function of the corresponding computational cost C_{TAMS} . For these particular dynamics, the TAMS is unable to efficiently sample rare trajectories associated to drag fluctuations higher than the largest fluctuation in the initial ensemble.

resampled trajectories never exceeds the amplitude of the highest maximum in the initial set of trajectories.

Figure ?? can be better understood by visualising the drag signals corresponding to the last resampled trajectories. Figure ?? displays the ensemble of trajectories after 181 iterations of the TAMS resampling procedure (see figure ??). One can see that, ultimately, trajectories are resampled from one unique initial trajectory. This phenomenon can be explained as follows. Because the flow dynamics have a memory, it takes a certain time before a resampled trajectory separate from its parent. In our case, this memory originates from the deterministic nature of the dynamics, which implies that two trajectories starting from infinitesimally close starting points will overlap over a timescale called the «Lyapunov timescale» [?]. However, we observe that, in our particular case, the duration of extreme drag fluctuations is typically 5 times smaller than the Lyapunov timescale of the drag process. This means that the resampling of a trajectory based on the value of the drag at t^* cannot lead to larger drag values for $t^* \leq t \leq \tau_L$. For $t > \tau_L$, the drag process has no memory of the drag fluctuations on which the resampling was based. As a consequence, the probability of observing an extreme fluctuation over $\tau_l \leq t \leq T_a$ is very low. The difference between the typical duration of drag fluctuations τ_c and the Lyapunov timescale τ_0 is linked to the convective nature of the flow. As described in section III B, extreme fluctuations of

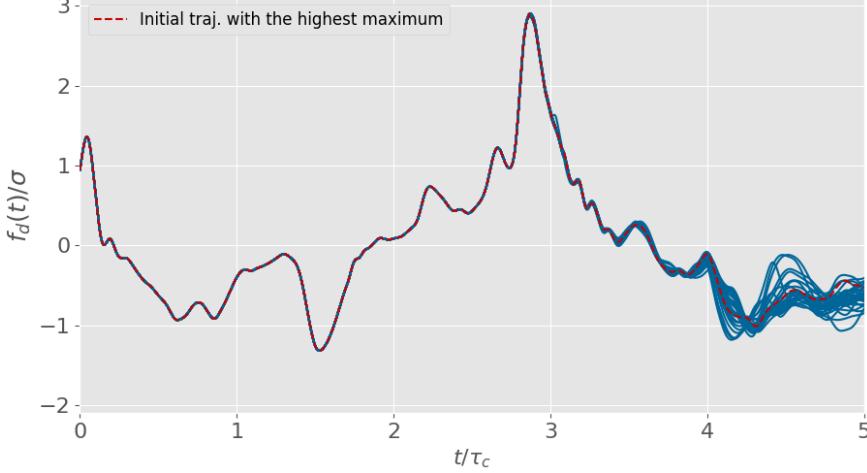


Figure 16: Ensemble of trajectories after 181 iterations of the selection-mutation procedure. In this experiment, the TAMS is used with the instantaneous drag f_d as a score function. Trajectories have a duration $T_a = 5\tau_c$ and their number is $N_c = 32$. The TAMS fails to efficiently sample rare trajectories, because all trajectories are ultimately resampled from the trajectory displaying the highest maximum in the initial ensemble.

the instantaneous drag f_d have a lifetime of roughly τ_c . Fluid structures responsible for a particular value of the drag on the obstacle at a time t^* are swept away by the mean flow over a timescale τ_c , shorter than the time it takes for trajectories to separate, *i.e* the Lyapunov timescale.

This result suggest that the TAMS—and AMS—cannot be expected to efficiently sample rare trajectories for dynamics with such a phenomenology. Note that, because of the complexity of the dynamics, we were limited to a rather low number of trajectories, relatively to previous use cases of these algorithms [?]. Targeting events with a fixed return time, one can show that the computational cost of a TAMS experiment roughly grows linearly with the number of trajectories N [?]. As a result, multiplying the number of trajectories by a factor 10 or 100 would require important computational resources as well as an efficient parallel implementation of the TAMS. Despite the relative simplicity of the TAMS algorithm, the latter implementation is difficult to achieve [?]. In any cases, from the observations made in this section, we do not expect that increasing the number of trajectories is able to better the efficiency of the sampling for such dynamics.

A promising perspective is to adapt the TAMS to dynamics for which the Lyapunov time is larger than the typical duration of extreme events, such as the convective flow involved

in this work. In the original TAMS, at later stages of the algorithm, new trajectories are resampled starting from points very close to the maximum of the parent trajectory. Resampled trajectories have no chance to display a significantly higher maximum as they overlap their parent trajectory over the Lyapunov timescale. As an example, the resampling procedure could be modified so that new trajectories are resampled roughly a Lyapunov time *before* the maximum is attained. In this way, it gives the opportunity for new events, associated to a higher fluctuations, to occur. However, such modification(s) of the resampling procedure must be implemented with care, so as to preserve the mathematical properties of the AMS and TAMS.

B. Extreme time-averaged drag forces with the Giardina-Kurchan-Tailleur-Lecomte algorithm

We now turn to the sampling of extreme fluctuations of the time-averaged drag F_T . The TAMS algorithm could be used in the same way as described in section IV A for the instantaneous drag, choosing F_T itself as a score function. However, this leads to the same problem: the resampling fails to efficiently generate new, higher fluctuations of F_T .

For time-averaged observables, one can make use of a different algorithm: the Giardina-Kurchan-Tailleur-Lecomte (GKTL) algorithm [? ? ?]. It was originally designed to compute large deviations rate functions of time-averaged dynamical observables, that describe the probability density of both typical of rare events in the limit of large averaging time [?]. By contrast with the AMS and TAMS, the GKTL algorithm does not follow the strategy of splitting algorithms. Instead, it implements *importance sampling*, *i.e.* sampling a modified distribution that is biased towards the rare events of interest. The idea of importance sampling is very general and was used in many different contexts (see e.g. [? ?] and the general references [? ?]). The GKTL algorithm performs importance sampling in the space of trajectories, which is relevant for out-of-equilibrium systems.

1. The GKTL algorithm

Similarly to the TAMS algorithm, the GKTL relies on the simulation of a ensemble of trajectories. In this case however, they interact dynamically: at regular time intervals, some

members of the ensemble are killed and some are cloned according to a weight which depends on the history of the replica. The weights are chosen such that, after several iterations of the algorithm, generated trajectories are distributed according to a probability distribution that is biased in favour of trajectories with large values of the time-averaged observable of interest. The GKTL algorithm belong to a family of algorithms known as «Go with the winners» [? ?]. Similar ideas have been applied in a wide range of fields over the past 50 years, under different names, depending of the specific application domain [?]. 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 [?]. Over the past ten years, it has successfully been applied to investigate rare events in both stochastic [? ? ? ?] and deterministic systems [? ?].

The GKTL algorithm consists in simulation an ensemble of N trajectories $\{\mathbf{x}_n(t)\}_{1 \leq n \leq N}$ starting from independant random initial conditions. Similarly to section IV A, the total integration time of the trajectories is denoted T_a . We consider an observable of interest $A(\mathbf{x}(t))$ and a cloning period τ . At times $t_i = i\tau$ (with $i = 1, 2, \dots, T_a/\tau$) we assign to each trajectory n a weight W_n^i defined as

$$W_n^i = \frac{e^{k \int_{t_{i-1}}^{t_i} A(X_n(t)) dt}}{R_i} \quad \text{with} \quad R_i = \frac{1}{N} \sum_{n=1}^N e^{k \int_{t_{i-1}}^{t_i} A(X_n(t)) dt}.. \quad (12)$$

For each trajectory $\{\mathbf{x}_n\}_{0 \leq t_i}$, a random number of copies of the trajectory are generated. This number is on average proportional to the weight W_n^i , such that the total number of trajectories produced at each event is equal to N . For deterministic systems, a small perturbation is introduced for clones at times t_i , so that clones separate from their parent. The parameter k sets the amplitude of the typical fluctuations within the biased ensemble of trajectory sampled by the algorithm. A higher value of k will give a higher weight to higher fluctuations, thus driving the sampling to more extreme events.

Let us denote formally $\mathbb{P}_0 (\{X(t)\}_{0 \leq t \leq T_a} = \{x(t)\}_{0 \leq t \leq T_a})$ the probability to observe a trajectory $\{x(t)\}_{0 \leq t \leq T_a}$ in the model, and $\mathbb{P}_k (\{X(t)\}_{0 \leq t \leq T_a} = \{x(t)\}_{0 \leq t \leq T_a})$ the probability to observe the same trajectory with the algorithm. By construction of the algorithm through

the weights (12), we have

$$\mathbb{P}_k \left(\{X(t)\}_{0 \leq t \leq T_a} = \{x(t)\}_{0 \leq t \leq T_a} \right) \underset{N \rightarrow \infty}{\sim} \frac{e^{k \int_0^{T_a} A(x(t)) dt}}{Z(k, T_a)} \mathbb{P}_0 \left(\{X(t)\}_{0 \leq t \leq T_a} = \{x(t)\}_{0 \leq t \leq T_a} \right). \quad (13)$$

where the normalisation factor is given by $Z(k, T_a) = \mathbb{E}_0 \left[e^{k \int_0^{T_a} A(X(t)) dt} \right]$, denoting by \mathbb{E}_0 the expectation value with respect to \mathbb{P}_0 , and $\underset{N \rightarrow \infty}{\sim}$ means that this is true only asymptotically for large N . The typical error is of order $1/\sqrt{N}$ when evaluating averages over observables. Equation (13) is obtained by assuming the mean field approximation

$$R_1 = \frac{1}{N} \sum_{n=1}^N e^{k \int_0^{t_1} A(X_n(t)) dt} \underset{N \rightarrow \infty}{\sim} Z(k, t_1) = \mathbb{E}_0 \left[e^{k \int_0^{t_1} A(X(t)) dt} \right], \quad (14)$$

which, by induction, and using a formula similar to (14) at each step of the induction, leads to [? ?]:

$$\prod_{i=1}^{T_a/\tau} R_i \underset{N \rightarrow \infty}{\sim} Z(k, T_a) = \mathbb{E}_0 \left[e^{k \int_0^{T_a} A(X(t)) dt} \right]. \quad (15)$$

The validity of the mean field approximation and the fact that the typical relative error due to this approximation is of order $1/\sqrt{N}$ has been proven [? ?] to be true for a family of rare event algorithms including the one adopted in this paper.

Formula (13) is valid only for times T_a that are integer multiples of the resampling time τ . The killed trajectories have to be discarded from the statistics. Starting from the final N trajectories at time T_a , one goes backwards in time through the selection events attaching to each piece of trajectory its ancestor. In this way one obtains an effective ensemble of N trajectories from time 0 to time T_a , distributed according to \mathbb{P}_k . All trajectories reconstructed in this way are real solutions of the model: we have not modified the dynamics, but only sampled trajectories according to the distribution \mathbb{P}_k rather than according to the distribution \mathbb{P}_0 .

The sampled trajectories 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 . This is done using the backward reconstructed trajectories and inverting formula (13). For more details concerning the implementation of the algorithm and the corresponding data analysis, see ?? As an example, say one wants to estimate the expectation value of an observable $O \left(\{X(t)\}_{0 \leq t \leq T_a} \right)$.

An estimator is then given by

$$\mathbb{E}_0 [O(\{X(t)\}_{0 \leq t \leq T_a})] \underset{N \rightarrow \infty}{\sim} \frac{1}{N} \sum_{n=1}^N O(\{X_n(t)\}_{0 \leq t \leq T_a}) e^{-k \int_0^{T_a} A(X_n(t)) dt} e^{T_a \lambda(k, T_a)}, \quad (16)$$

where the X_n are the N backward reconstructed trajectories. Events that can be considered rare with respect to \mathbb{P}_0 are oversampled following \mathbb{P}_k . As a result, the effective ensemble of trajectories generated by the GKTL algorithm contains a larger amount of such events, and empirical estimators such as (16) have a dramatically lower statistical error.

2. Application to extreme fluctuations of the time-averaged drag

We now discuss the application of the GKTL algorithm to the flow dynamics introduced in section II. The objective is to sample trajectories displaying extreme fluctuations of the *time-averaged* drag (8). Consistently with section III C, we choose the averaging window $T = 10\tau_c$.

The computational cost C_{gktl} of a GKTL experiment is determined by both the duration of the trajectories T_a and the number of trajectories N_c : $C_{gktl} = N_c \times T_a$. In the following series of experiments we set $T_a = T = 10\tau_c$. The sampling from the GKTL algorithm will be compared to the sampling from the control run, *i.e.* a direct simulation of the flow over $T_{tot} = 10^6\tau_c$. We set $N_c = 16384$, leading to $C_{gktl} \approx 1.6 \times 10^5\tau_c$. (**Plus vraiment clair pour moi pourquoi je n'ai pas utilisé 10X plus de clones, pour matcher le coût numérique de la simulation de contrôle...**) Lastly, one must set the cloning period τ . On the one hand, too small a cloning period may result in loss of information: clones do not separate from their parent over a cloning period. On the other hand, choosing $\tau \gg \tau_c$ may result in too few cloning stages to achieve importance sampling. A rule of thumb is to set $\tau \approx \tau_c$. In the following, we set $\tau = \tau_c/2$.

We performed three experiments, corresponding to three different values of the parameter k . In general the value of k must be set empirically, unless a *large deviation regime* $T_a \rightarrow \infty$ is verified [?]. In the latter case, the parameter k can be related to the typical amplitude of the sampled fluctuations through the Gartner-Ellis theorem [?].

Figure ?? illustrates importance sampling achieved by means of the GKTL algorithm for the time-averaged drag F_T with $T = T_a = 10\tau_c$. For each of the three experiments, it

displays the estimate of the biased PDF

$$\rho_k(F) = \mathbb{E}_{\mathcal{P}_k} [\delta(F_{T_a}[\{\mathbf{x}\}_{0 \leq t \leq T_a}] - F)] \approx \frac{1}{N_c} \sum_{j=1}^{N_c} \delta(F_{T_a}^{(j)}[\{\mathbf{x}\}_{0 \leq t \leq T_a}] - F), \quad (17)$$

computed over the set of $N = 16384$ trajectories sampled by the algorithm. Moreover, figure ?? displays an estimate of the «unbiased» PDF \mathcal{P}_0 describing the time-averaged drag process. Note that it corresponds to equation (17) with $k = 0$.

For a fixed number of trajectories N , there is a limited range of k values for which the algorithm is able to yield an accurate estimate of ρ_k . Setting k to too high a value causes harmful effects resulting from the finiteness of the ensemble of trajectories: trajectories overlap over most of their duration. This effect is illustrated in figure ??, where the histogram corresponding to the largest value of k is peaked around a single value. This is a consequence of the fact that most of the trajectories in the biased ensemble overlap, therefore leading to the same value for the averaged drag F_T . See [? ? ?] for further details about the scaling of the required population size as k is increased.

Unsurprisingly, sampling larger fluctuations requires a larger population size, and therefore a larger computational cost. Even though, table I shows that the GKTL algorithm is able to sample trajectories corresponding to values of the time-averaged drag that could not be sampled through a direct sampling with equal computational cost. Table I displays the number of trajectories that correspond to a time-average $F_{T_a} \geq a$, for several values of a . As a comparison, it also displays the average number of such fluctuations in a control timeseries of duration C_{gktl} , divided in $N_c = C_{gktl}/T_a$ events. This number is computed based on the assumption that the PDF describing the random variable F_T with $T = 10\tau_c$ is Gaussian. This hypothesis is motivated by figure 17. While this assumption does not hold for $F_T \geq 4\sigma_T$, it may still provide numbers relevant to the comparison with the algorithm.

In section III C we observed that extremely large values of the time-averaged drag can either result from a succession of rather typical—however all positive—fluctuations of the instantaneous drag f_d , or from a small number of extremely large fluctuations of the instantaneous drag. In the latter case, each fluctuation typically lasts a correlation time τ_c . In this context, the GKTL algorithm is particularly suited to sample trajectories that display a succession of these fluctuations. (**conclusion à développer.**)

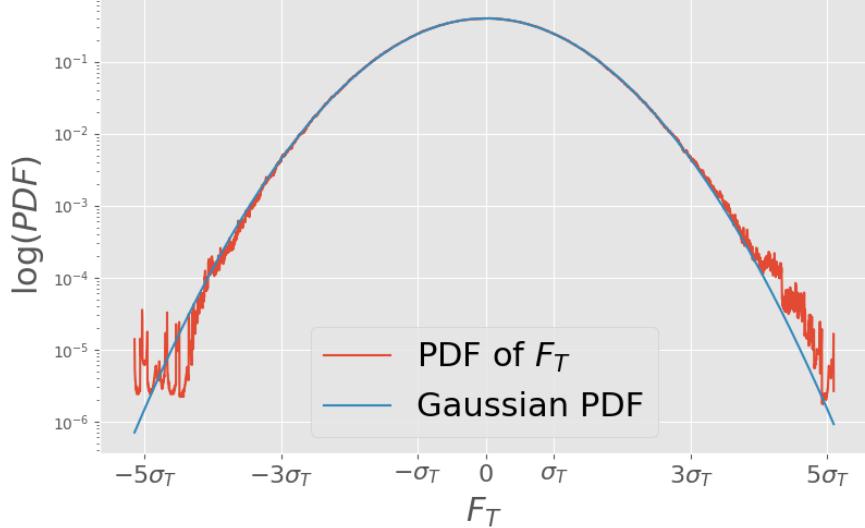


Figure 17: PDF of average drag over $10\tau_c$

	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
\mathcal{N}_{direct}	2599.408	372.738	22.117	0.519	0.005

Table I: Number of independent corresponding to fluctuations $F_{T_a} \geq a$ with $a=\sigma, 2\sigma\dots$ in the biased ensemble, for the three experiments $k = 0.02, 0.025, 0.03$. Parameters of the GKTL experiments are $N_c = 16384$, $T_a = 10\tau_c$, $\tau = \tau_c/2$. Within the biased ensemble, trajectories are considered distinct if they overlap for less than 50% of their duration. In addition, \mathcal{N}_{direct} is the average number of fluctuations $F'_{T_a} \geq a$ one can expect from N_c independent realisations of F_{T_a} , *i.e.* with $k = 0$. This number is based on the fact that the PDF of F_{T_a} is Gaussian.

V. CONCLUSION

Figure 18: **(a)**: Importance sampling for the time-averaged drag acting on the obstacle embedded in the test flow. The drag is averaged over $T = 10\tau_c$. The GKTL algorithm is used with $N_c = 16384$ copies and $T_a = T = 10\tau_c$ for three increasing values of the bias k . Additionally, the cloning period is $\tau = \tau_c/2$, resulting in $M = 20$ iterations of the evolution/selection procedure. Note that in this particular example the last cloning stage at t_{19} is not performed. As k is increased, the sampled PDF is shifted towards extreme values. Along the X-axis we display the fluctuation $F'_T = F_T - \overline{F_T}$, where the average $\overline{F_T}$ and standard deviation σ_T have been computed over a timeseries spanning $10^6\tau_c$. One can see that the distribution for $k = 0.03$ is peaked for a particular value of the fluctuation. This indicates that a significant part of the sampled trajectories actually overlap over the major part of their history. As a result, the average drag only differ by a small amount.