Gravitational Turbulence: the Small-Scale Limit of the Cold-Dark-Matter Power Spectrum

Yonadav Barry Ginat,[1,2] Michael L. Nastac,[1,3] Robert J. Ewart,[1,4]
Sara Konrad,[5] Matthias Bartelmann,[5] and Alexander A. Schekochihin[1,6]

[1] *Rudolf Peierls Centre for Theoretical Physics, University of Oxford, Parks Road, Oxford, OX1 3PU, United Kingdom*
[2] *New College, Holywell Street, Oxford, OX1 3BN, United Kingdom*[*]
[3] *St. John's College, St. Giles, Oxford, OX1 3JP, United Kingdom*
[4] *Balliol College, Broad Street, Oxford, OX1 3BJ, United Kingdom*
[5] *Institute for Theoretical Physics, Heidelberg University, Germany*
[6] *Merton College, Merton Street, Oxford, OX1 4JD, United Kingdom*

The matter power spectrum, $P(k)$, is one of the fundamental quantities in the study of large-scale structure in cosmology. Here, we study its small-scale asymptotic limit, and show that for cold dark matter in d spatial dimensions, $P(k)$ has a universal k^{-d} asymptotic scaling with the wave-number k, for $k \gg k_{\rm nl}$, where $k_{\rm nl}$ denotes the scale at which non-linearities in gravitational interactions become important. We propose a theoretical explanation for this scaling, based on a non-perturbative analysis of the system's phase-space structure. Gravitational collapse is shown to drive a turbulent phase-space cascade of the quadratic Casimir invariant, where the linear and non-linear time scales are balanced. A parallel is drawn to Batchelor turbulence in hydrodynamics, where here large scales mix smaller ones via tides. The k^{-d} scaling is also derived by expressing $P(k)$ as a phase-space integral in the framework of kinetic field theory, which is analysed by the saddle-point method; the dominant critical points of this integral are precisely those where the time scales are balanced. The coldness of the dark-matter distribution function—its non-vanishing only on a d-dimensional sub-manifold of phase-space—underpins both approaches. The theory is accompanied by 1D Vlasov–Poisson simulations, which confirm it.

I. INTRODUCTION

One of the basic observable quantities in the study of the large-scale structure of the Universe is the two-point correlation function of the over-density field, whose Fourier transform is the power spectrum, $P(k,t)$ [1, 2]. The two-point correlation function is of fundamental importance, for it allows us to probe theories of the early Universe, dark matter, inflation, and to study gravity [1, 3–5]. In this paper, we will explore the small-scale asymptotic behaviour of $P(k)$ (we omit the explicit time dependence when it is not confusing to do so), in the limit $k \gg k_{\rm nl}$, where the (inverse) non-linear scale $k_{\rm nl}$ is defined by $\int_{k_{\rm nl}}^{\infty} P_{\rm lin}(k) k^2 {\rm d}k = 2\pi^2\delta_c^2$, with δ_c denoting the spherical-collapse threshold [6]. The small-scale limit of $P(k)$ is theoretically important for the understanding of the gravitational N-body problem in the large-N limit [7], and the formation of large-scale structure, non-linear clustering and self-similarity [1, 8], but also for the understanding of the nature of dark matter and gravitational back-reaction of small scales on large ones—both relativistic [9] and in the context of the effective field theory of large-scale structure [10] (for reviews and references see, e.g., [6, 11, 12]) and the general bias expansion [3]. Even before modifying gravity, it is important to know what non-linear phenomena occur in the standard theory.

Data from cosmological dark-matter-only simulations are consistent with $P(k,t)$ developing a k^{-d} tail on small scales [e.g., 13, fig. 6] in 3D, and in 1D, as shown in refs. [14, figure 6] or [15, figure 1] (cf. [16]). The emergence of a power-law tail—and the simplicity of its exponent—hint that a fundamental physical reason for it must exist, ultimately stemming from the nature of the gravitational interaction of cold dark matter. Here, we will describe the mechanism that produces this asymptotic scaling, by studying the mass distribution in the velocity space as well as in position space.

We restrict ourselves to the strictly collisionless case where the particle mass $m \to 0$, while the total particle number $N \to \infty$, so that $M \equiv Nm$ remains fixed (and so does the volume). In this limit, the phase-space distribution of particles is well described by the Vlasov equation [1, 14, 17–20]:[1]

$$\frac{\partial f}{\partial \eta} + \mathbf{v} \cdot \frac{\partial f}{\partial \mathbf{x}} + \mathbf{g} \cdot \frac{\partial f}{\partial \mathbf{v}} = 0, \tag{1}$$

where f is the distribution function (the one-point probability density in phase-space), η is the conformal time, defined by ${\rm d}t = a{\rm d}\eta$, where a is the scale factor of the background (which is taken to be a Friedmann–Lemaître–Robertson–Walker space-time), $\mathbf{x}$ and $\mathbf{v}$ are

* yb.ginat@physics.ox.ac.uk

[1] The collisionless Vlasov equation, of course, ignores dissipation via collisions (or equivalently, finite-N effects). However, we will find that due to turbulence, such dissipation will inevitably be accessed; this point will be discussed further in appendix A 2.

the co-moving position and velocity, and the gravitational field $\mathbf{g}$ is self-consistently derived from Poisson's equation

$$\boldsymbol{\nabla}\cdot\mathbf{g}\equiv-\nabla^2\Phi=-4\pi Ga^2\left[\int f(\mathbf{x},\mathbf{v},\eta)\mathrm{d}^dv-\overline{\rho}_m\right], \quad (2)$$

where $\overline{\rho}_m$ is the mean background matter density and Φ is the gravitational potential. This system of equations applies well on scales much smaller than the horizon, for particles much slower than the speed of light, c. Henceforth, we will work in general spatial dimensions $d\in\{1,2,3\}$, to enable comparison between our theory and $d=1$ simulations.

Dark matter is assumed to be cold initially, with a thermal velocity $v_{\rm th}\to 0$, so that the initial gravitational potential energy is much larger than the initial thermal energy (this assumption is excellent for our Universe [4, 21–26]). Clearly, despite being cold, dark matter is inherently kinetic, and cannot be described by fluid equations adequately on non-linear scales, because these equations cease to be valid after streams cross [20, 27, 28].

Below we will derive the $k\gg k_{\rm nl}$ limit of $P(k)$ in two complementary ways. First, we will study the problem in phase-space, and show that the small-scale asymptotics arise when the time scales involved in the Vlasov equation balance with each other in a particular way, to be explained below. Two ingredients will comprise this argument: the balance of time scales, and the conservation of the second Casimir invariant [29–33]

$$C_2\equiv\iint f^2\mathrm{d}^dx\mathrm{d}^dv. \quad (3)$$

This invariant is sometimes referred to as 'enstrophy', 'phasestrophy', or 'f-strophy'. The Vlasov equation conserves an infinite number of phase-space invariants—not only C_2—but we will use C_2 here because it is directly related to the power spectrum (*vide infra*). We will show that the spectrum can be predicted if small-scale structure is understood as resulting from a turbulent cascade of C_2 from large scales to small ones.

For the second approach, we use non-perturbative kinetic field theory (KFT) for cosmic structure formation (for a review, see [34]). Here, $P(k)$ is expressed as an integral over the initial particle positions and velocities (weighted by the initial-condition probability distribution) of the characteristic function of the displacement field; this integral will turn out to have an explicit k dependence, and we will utilise this to perform an asymptotic saddle-point analysis.

The two approaches complement each other, both relying on the same assumptions, but highlighting their rôles in different ways. We remark that using phase-space expressions ensures that the validity of our results extends beyond the bounds of configuration-space-based approaches, such as Lagrangian [16, 35, 36] or Eulerian techniques [27]; in particular, it is regular at stream-crossing, and accounts for free streaming automatically. Indeed, as we already mentioned above, the inherently kinetic phenomenon of multi-scale structure of the distribution of dark matter, developing by virtue of strongly non-linear interactions, suggests that a type of turbulence in phase-space is involved. In turbulent phase-space dynamics, there is a flux of C_2 from large scales to small ones [32], and we will show below that C_2 cascades to smaller scales by gravitational collapse too. As early as [36], it was realised that the phenomena of gravitation and turbulence might be linked—here we make the analogy precise and characterise this gravitational turbulence.

The rest of this paper is organised as follows: in §II, we formulate the first approach, based on the Vlasov–Poisson system, and show how the concept of a 'phase-space cascade' can be used to derive the small-scale asymptotics of the phase-space power spectrum, from which $P(k)$ may be computed. In §III, as promised above, we derive the same asymptotic scaling of $P(k)$ again, but via a saddle-point analysis of an integral expression for $P(k)$. We test our theory by comparing it with numerical Vlasov–Poisson simulations throughout the paper. Our conclusions are discussed in §IV and summarised in §V.

II. PHASE-SPACE TURBULENCE

We start by describing the initial condition in §II A—a cold stream or a collection of streams—and show what its evolution looks like, and then derive the equation that governs the phase-space Fourier transform of f in §II B. We will then analyse the time scales involved in the Vlasov–Poisson system in §II C, and show in §II D that a turbulent flux of C_2 to smaller scales characterises the dynamics. To do that, we will derive a transport equation for the integrand of C_2 with a source term; in §II E this source will be found to receive contributions from all larger scales, because of the Jeans instability [37]. This in turn will allow us to find the asymptotic scaling of $P(k)$ in §II F. We will present simulation results throughout this section, to test the theory.

A. Cold streams

We assume that the system has an initial condition consisting of a superposition of streams, each one of the form:

$$f(\eta=0,\mathbf{x},\mathbf{v})=f_{\rm in}(\mathbf{x},\mathbf{v})\equiv\frac{\rho_{\rm in}(\mathbf{x})}{(2\pi v_{\rm th}^2)^{d/2}}\mathrm{e}^{-\frac{[\mathbf{v}-\mathbf{u}_{\rm in}(\mathbf{x})]^2}{2v_{\rm th}^2}}, \quad (4)$$

where $v_{\rm th}^2\ll\min\left\{u_{\rm in}^2,\int|\Phi_{\rm in}|\,f_{\rm in}\mathrm{d}^dx\mathrm{d}^dv\right\}$, where $\Phi_{\rm in}$ is derived from $\rho_{\rm in}$ via equation (2); that is, the initial condition is very cold—the thermal energy is negligible in comparison with the gravitational potential energy of the system or the kinetic energy of mean flows.

In equation (4), $\rho_{\rm in}$ and $\mathbf{u}_{\rm in}$ are typically Gaussian random fields [1, 3], and $\mathbf{u}_{\rm in}$ is a gradient flow [2]. We also choose the co-ordinates so that $\int \rho(\mathbf{x})\mathbf{u}(\mathbf{x})\mathrm{d}^d x = 0$, where $\rho(\mathbf{x})$ is the density and $\rho\mathbf{u} \equiv \int \mathbf{v} f \mathrm{d}^d v$. The distribution (4) is essentially a single stream, and, in fact, its Maxwellian shape may be approximated by a Dirac delta-function. The thermal speed $v_{\rm th}$ is chosen to be the smallest velocity scale in the problem, so the entire analysis of this paper focuses on the limit where $s v_{\rm th} \ll 1$, where s is the Fourier conjugate to velocity (see §II B). The distribution function remains a collection of streams as it evolves in time, by Liouville's theorem [e.g. 38]: locally almost everywhere in phase-space (in a sufficiently small neighbourhood of almost any point $(\mathbf{x}_0, \mathbf{v}_0)$ where $f \neq 0$), it can be written as $\rho_{\mathbf{x}_0,\mathbf{v}_0}(\mathbf{x},\eta)\delta^{\rm D}\left(\mathbf{v} - \mathbf{u}_{\mathbf{x}_0,\mathbf{v}_0}(\mathbf{x},\eta)\right) + O(v_{\rm th})$, where $\delta^{\rm D}$ is the Dirac delta function; it remains so as long as collisions or finite-N effects may be ignored. The local functions $\rho_{\mathbf{x}_0,\mathbf{v}_0}(\mathbf{x},\eta)$ and $\mathbf{u}_{\mathbf{x}_0,\mathbf{v}_0}(\mathbf{x},\eta)$ might in general differ from the density $\rho(\mathbf{x},\eta)$ and the mean velocity $\mathbf{u}(\mathbf{x},\eta)$.

To study the evolution of the initial condition (4), we have conducted a suite of Vlasov–Poisson simulations in 1D on a Minkowski background ($a(\eta) = 1$), using the Gkeyll code [39], which is originally a Vlasov solver for kinetic plasmas; by setting the vacuum permittivity to $\varepsilon_0 < 0$, we can use the code to study gravity as opposed to electrostatics. Details on the numerical method and simulation set-up may be found in appendix A. The simulation's units are chosen so that $\tau_0^{-2} \equiv 4\pi G M/L = 1$ (unit of time),[2] $k_0 \equiv 2\pi/L = 1$ (unit of length) and $v_0 \equiv (k_0\tau_0)^{-1} = 1$ (unit of speed), where L is the length of the simulation 'box' (with periodic boundary conditions) and M is the total mass in the box.

The time evolution of a cold system of three streams—three copies of equation (4)—is displayed in figure 1. This figure shows that each stream is distorted quickly, by rotating and twisting in phase-space. Evidently, this motion generates small-scale structure, which we consider to be a type of turbulence in phase-space. As usual, in order to characterise the turbulence and its spectrum of fluctuations, one must identify an invariant quantity that cascades to small scales, what the cascade's time scale is, and also the flux of that invariant quantity [40]. We will do so in §§II D, II C, and II E, respectively.

That the turbulence is in phase-space and not merely in position space is clear: already at $t = 6\tau_0$, the system can no longer be described as three streams almost anywhere, so standard cosmological perturbation theory would already be inadequate, and the root-mean-square (rms) peculiar velocity $v_{\rm rms} \equiv \left[\int \mathrm{d}v\, (v - u(x))^2 f(x,v)/\rho(x)\right]^{1/2}$ is much larger than $v_{\rm th}$, and is of order v_0. However, in phase-space, we see visually that the topology of three single lines is preserved.

The fact that the system is a collection of streams implies that *locally in phase-space*, one can describe each stream with fluid equations: inserting $\rho_{\mathbf{x}_0,\mathbf{v}_0}(\mathbf{x},\eta)\delta^{\rm D}\left(\mathbf{v} - \mathbf{u}_{\mathbf{x}_0,\mathbf{v}_0}(\mathbf{x},\eta)\right)$ into the Vlasov equation gives the continuity and Euler equations for $\rho_{\mathbf{x}_0,\mathbf{v}_0}$ and $\mathbf{u}_{\mathbf{x}_0,\mathbf{v}_0}$, respectively, and the divergence of the latter yields the Raychaudhuri equation [41], which describes gravitational collapse under the Jeans instability [20, 35, 37, 42–48]. Written in the frame of reference that moves with a given stream, it reads

$$\frac{\mathrm{d}\theta}{\mathrm{d}\eta} + \mathcal{H}\theta + \frac{\theta^2}{3} + \sigma^{ij}\sigma_{ij} = \boldsymbol{\nabla}\cdot\mathbf{g} + \frac{\Sigma}{\rho_{\mathbf{x}_0,\mathbf{v}_0}} + 2\omega_i\omega^i, \quad (5)$$

where $\mathcal{H}$ is the conformal Hubble constant, $\theta \equiv \boldsymbol{\nabla}\cdot\mathbf{u}_{\mathbf{x}_0,\mathbf{v}_0}$ is the stream's divergence, $\boldsymbol{\omega} \equiv \boldsymbol{\nabla}\times\mathbf{u}_{\mathbf{x}_0,\mathbf{v}_0}/2$ is its vorticity, $\sigma_{ij} \equiv \left[\partial_i(u_{\mathbf{x}_0,\mathbf{v}_0})_j + \partial_j(u_{\mathbf{x}_0,\mathbf{v}_0})_i - 2\theta\delta_{ij}/3\right]/2$ is its shear, $\mathrm{d}/\mathrm{d}\eta$ is the Lagrangian derivative along the stream, $\Sigma/\rho_{\mathbf{x}_0,\mathbf{v}_0}$ is a pressure-related term that is $O(v_{\rm th}^2)$. Below, $\mathbf{g}$ is the gravitational field felt by the stream. We will use equation (5) to estimate the time scale $\tau_{\rm J}$ of gravitational collapse.

B. Batchelor approximation

We need to characterise a turbulence, which is inherently a multi-scale process, so it is more convenient to study equations (1)-(2) in Fourier space.

1. *Fourier transform*

Let is define the Fourier transform (marked by a circumflex) as follows:

$$\hat{f}(\mathbf{k},\mathbf{s}) \equiv \iint f(\mathbf{x},\mathbf{v})\mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\mathbf{x} - \mathrm{i}\mathbf{s}\cdot\mathbf{v}}\mathrm{d}^d x \mathrm{d}^d v, \quad (6)$$

and similarly for all other functions of $(\mathbf{x},\mathbf{v})$. For $d > 1$, we denote $k \equiv |\mathbf{k}|$, $s \equiv |\mathbf{s}|$, *etc.* Under this Fourier transform, the Vlasov–Poisson system (1–2) becomes

$$\frac{\partial \hat{f}}{\partial \eta} + \mathbf{k}\cdot\frac{\partial \hat{f}}{\partial \mathbf{s}} + \mathrm{i}\mathbf{s}\cdot\int \frac{\mathrm{d}^d k'}{(2\pi)^d}\hat{\mathbf{g}}(\mathbf{k}')\hat{f}(\mathbf{k}-\mathbf{k}',\mathbf{s}) = 0, \quad (7)$$

$$k^2\hat{\Phi} = -4\pi G a^2 \hat{\rho}, \quad (8)$$

where $\hat{\rho}$ is the Fourier-transformed density. The second Casimir invariant, defined by equation (3), is given in Fourier space by Parseval's theorem:

$$C_2 = \frac{1}{(2\pi)^{2d}}\iint |\hat{f}|^2 \mathrm{d}^d k \mathrm{d}^d s. \quad (9)$$

The integrand $|\hat{f}^2|$ is directly related to the density power spectrum. Indeed, let the *phase-space power spectrum* be

$$\hat{F}(\mathbf{k},\mathbf{s}) \equiv \langle |\hat{f}|^2 \rangle, \quad (10)$$

[2] The Poisson equation (2) implies that G has different dimensions in 1D from 3D.

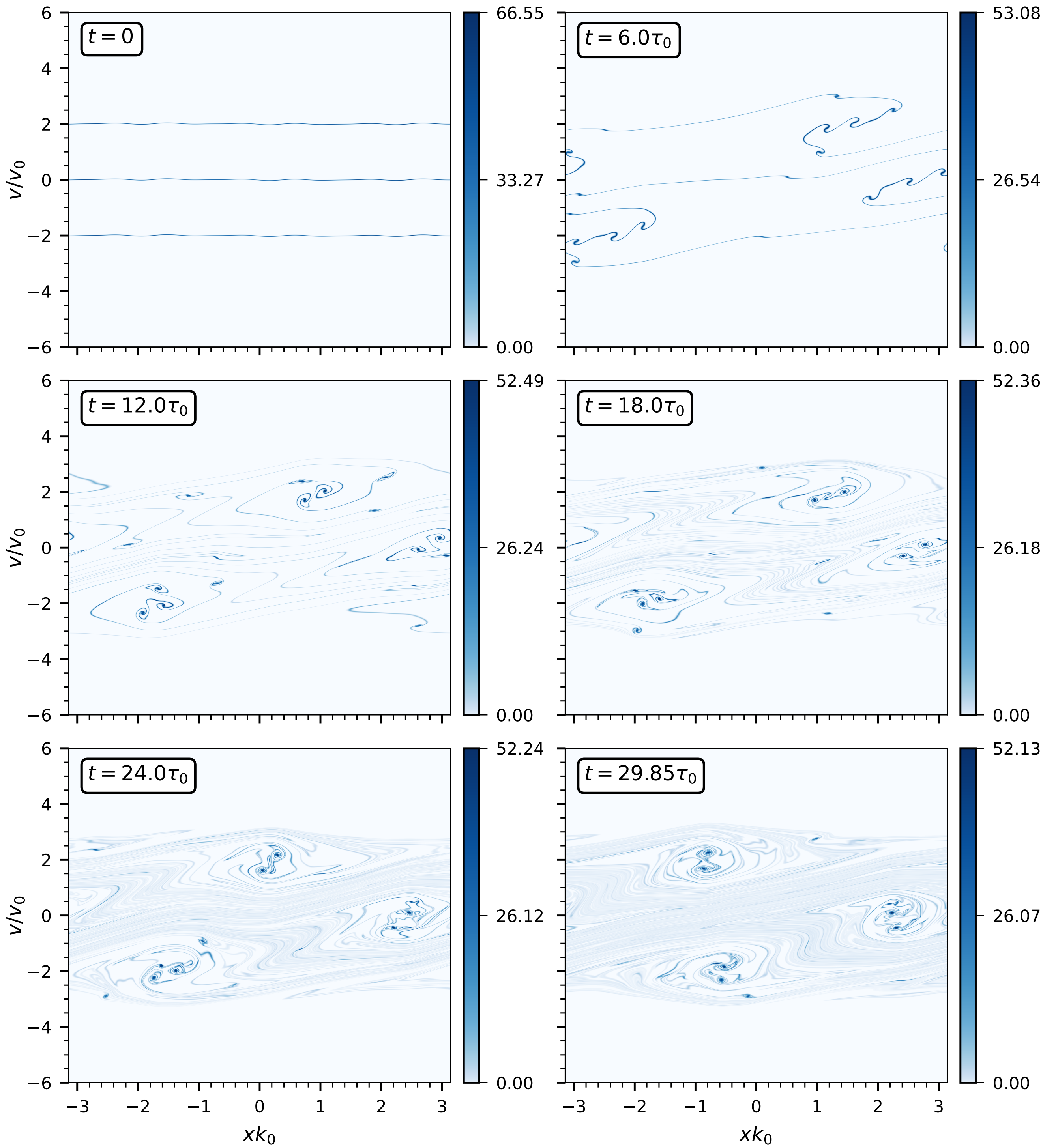

FIG. 1. Colour plots of the distribution function showing the time evolution of three cold streams. See text and appendix A for details. A video is available here.

where $\langle\cdot\rangle$ is an ensemble average over many random realisations of the initial conditions; then the density power spectrum is

$$P(k) \equiv \frac{1}{V} \langle \hat{\rho}(\mathbf{k}) \hat{\rho}^*(\mathbf{k}) \rangle = \frac{\hat{F}(\mathbf{k}, 0)}{V}, \tag{11}$$

where V is the spatial volume.

2. Flux of C_2

Let us now see how the integrand of the second Casimir invariant (9) evolves: multiplying equation (7) by $\hat{f}^*$ and taking the real part, we find

$$\frac{\partial|\hat{f}|^2}{\partial\eta}+\mathbf{k}\cdot\frac{\partial|\hat{f}|^2}{\partial\mathbf{s}}+\mathrm{i}\mathbf{s}\cdot\int\frac{\mathrm{d}^dk'}{(2\pi)^d}\left[\hat{\mathbf{g}}(\mathbf{k}')\hat{f}^*(\mathbf{k},\mathbf{s})\hat{f}(\mathbf{k}-\mathbf{k}',\mathbf{s})-\hat{\mathbf{g}}^*(\mathbf{k}')\hat{f}(\mathbf{k},\mathbf{s})\hat{f}^*(\mathbf{k}-\mathbf{k}',\mathbf{s})\right]=0. \tag{12}$$

Integrating equation (12) over all $\mathbf{k}$ and $\mathbf{s}$ yields the conservation of C_2, as it should.

Let δg_r be the amplitude of a typical fluctuation in $\mathbf{g}$ on scales $r=1/k$ and smaller, *viz.*,

$$\delta g_r^2\equiv\frac{2}{V}\int\frac{\mathrm{d}^dk'}{(2\pi)^d}\ \left\langle|\hat{\mathbf{g}}(\mathbf{k}')|^2\right\rangle\left(1-\mathrm{e}^{\mathrm{i}\mathbf{k}'\cdot\mathbf{r}}\right)\sim\frac{1}{V}\int_k^\infty\mathrm{d}^dk'\ \left\langle|\hat{\mathbf{g}}(\mathbf{k}')|^2\right\rangle. \tag{13}$$

We take $\mathbf{g}$ to be a sufficiently continuous field, so that for sufficiently small values of r,

$$\delta g_r\sim\kappa r^\lambda+\text{h.o.t.} \tag{14}$$

with Hölder exponent $\lambda\leq 1$, and κ a constant coefficient. A smooth gravitational field has $\lambda=1$, because one may Taylor-expand $\mathbf{g}$, meaning that the fluctuations are dominated by tidal forces (and κ has dimensions of $[\text{time}]^{-2}$).[3] We prove in §II F that $\lambda=1$ follows from Poisson's equation (2), in conjunction a balance of time scales (§II C) and a C_2 cascade, in $d\leq 3$ spatial dimensions, but to simplify the exposition we will take $\lambda=1$ now, *a priori*.

Let us now take the large-k limit of equation (12). Then, for a smooth gravitational field ($\lambda=1$), the last term on the left-hand side is dominated by two contributions: (i) $k'\ll k$ and (ii) $|\mathbf{k}-\mathbf{k}'|\ll k$. These are sometimes referred to as 'squeezed' triangles; case (i) is also known as the Batchelor limit [32, 49, 50]. Taylor-expanding $\hat{f}(\mathbf{k}-\mathbf{k}',\mathbf{s})$ in the Batchelor limit turns the non-linear term in (12) into

$$\begin{aligned}&\mathrm{i}\mathbf{s}\cdot\int\frac{\mathrm{d}^dk'}{(2\pi)^d}\left[\hat{\mathbf{g}}(\mathbf{k}')\hat{f}^*(\mathbf{k},\mathbf{s})\left(\hat{f}(\mathbf{k},\mathbf{s})-\frac{\partial\hat{f}}{\partial k^i}k'^i\right)-\hat{\mathbf{g}}^*(\mathbf{k}')\hat{f}(\mathbf{k},\mathbf{s})\left(\hat{f}^*(\mathbf{k},\mathbf{s})-\frac{\partial\hat{f}^*}{\partial k^i}k'^i\right)\right]+\text{h.o.t.}\\&\simeq-\mathrm{i}\int\frac{\mathrm{d}^dk'}{(2\pi)^d}s^j\left[\hat{g}_j(\mathbf{k}')k'^i\hat{f}^*(\mathbf{k},\mathbf{s})\frac{\partial\hat{f}}{\partial k^i}-\hat{g}_j^*(\mathbf{k}')k'^i\hat{f}(\mathbf{k},\mathbf{s})\frac{\partial\hat{f}^*}{\partial k^i}\right]\\&=-\mathrm{i}\int\frac{\mathrm{d}^dk'}{(2\pi)^d}s^j\hat{g}_j(\mathbf{k}')k'^i\left[\hat{f}^*(\mathbf{k},\mathbf{s})\frac{\partial\hat{f}}{\partial k^i}+\hat{f}(\mathbf{k},\mathbf{s})\frac{\partial\hat{f}^*}{\partial k^i}\right]\equiv-\mathrm{i}s^j\Phi^i_j\frac{\partial|\hat{f}|^2}{\partial k^i}.\end{aligned} \tag{15}$$

The transition from the first line to the second in (15) is valid because, in the centre-of-mass frame, the leading-order terms are proportional to $\int\mathrm{d}^dk'\hat{\mathbf{g}}(\mathbf{k}')=0$.

Thus, the non-linear term reduces to a tidal interaction: on a given (small) scale corresponding to a (large) wave-number k, the distribution function f is distorted by the gravitational field at the same (large) energy-containing scale—the matrix $\Phi^i_j=-\mathrm{i}\delta^{in}\partial_n\partial_j\Phi$ is the Hessian matrix of the gravitational potential, i.e., the tidal matrix.

3. Ensemble average

We now have, from equations (12) and (15),

$$\frac{\partial|\hat{f}|^2}{\partial\eta}+\mathbf{k}\cdot\frac{\partial|\hat{f}|^2}{\partial\mathbf{s}}-\mathrm{i}s^j\Phi^i_j\frac{\partial|\hat{f}|^2}{\partial k^i}=-(\text{ii}), \tag{16}$$

where (ii) represents the last term in equation (12) in the limit (ii), $|\mathbf{k}-\mathbf{k}'|\ll k$; upon substituting $\mathbf{k}''=\mathbf{k}-\mathbf{k}'$ in equation (12), one has

$$\begin{aligned}(\text{ii})&\simeq\mathrm{i}\mathbf{s}\cdot\int_{k''\ll k}\frac{\mathrm{d}^dk''}{(2\pi)^d}\left[\hat{\mathbf{g}}(\mathbf{k}-\mathbf{k}'')\hat{f}^*(\mathbf{k},\mathbf{s})\hat{f}(\mathbf{k}'',\mathbf{s})-\hat{\mathbf{g}}^*(\mathbf{k}-\mathbf{k}'')\hat{f}(\mathbf{k},\mathbf{s})\hat{f}^*(\mathbf{k}'',\mathbf{s})\right]+\text{h.o.t.}\\&\simeq\frac{\mathrm{i}\mathbf{s}}{V}\cdot\left[\hat{\mathbf{g}}(\mathbf{k})\hat{f}(0,\mathbf{s})\hat{f}^*(\mathbf{k},\mathbf{s})-\hat{\mathbf{g}}^*(\mathbf{k})\hat{f}^*(0,\mathbf{s})\hat{f}(\mathbf{k},\mathbf{s})\right]\simeq\mathrm{i}\mathbf{s}\cdot\left[\hat{\mathbf{g}}(\mathbf{k})\hat{\overline{f}}(\mathbf{s})\hat{f}^*(\mathbf{k},\mathbf{s})-\hat{\mathbf{g}}^*(\mathbf{k})\hat{\overline{f}}^*(\mathbf{s})\hat{f}(\mathbf{k},\mathbf{s})\right],\end{aligned} \tag{17}$$

---

[3] The case $\lambda>1$ is also smooth, but highly atypical, where the tidal forces vanish; while this could happen in isolated points with exactly zero over-density, we ignore it here.

where $\overline{f} \equiv \int f \mathrm{d}^d x/V$ is the volume average of f, and we have approximated the integral $\int_{k''\ll k} \mathrm{d}^d k'' \sim (2\pi)^d V^{-1}$ by the contribution from the smallest wave-number.

Equation (16) is a transport equation in Fourier space, with a source $-$(ii). Taking its average yields an evolution equation for the power spectrum (10):

$$\frac{\partial \hat{F}}{\partial \eta} + \mathbf{k}\cdot\frac{\partial \hat{F}}{\partial \mathbf{s}} - \mathrm{i}s^j \frac{\partial}{\partial k^i}\left\langle \Phi^i_j |\hat{f}|^2 \right\rangle = \hat{S}, \tag{18}$$

with the source

$$\hat{S} = -\iint \mathrm{d}^d x_1 \mathrm{d}^d v_1 \mathrm{d}^d x_2 \mathrm{d}^d v_2 \; \mathrm{e}^{\mathrm{i}\mathbf{k}\cdot(\mathbf{x}_1-\mathbf{x}_2)-\mathrm{i}\mathbf{s}\cdot(\mathbf{v}_1-\mathbf{v}_2)} \left\langle f(\mathbf{x}_1,\mathbf{v}_1)\mathbf{g}(\mathbf{x}_2)\cdot\frac{\partial \overline{f}(\mathbf{v}_2)}{\partial \mathbf{v}_2} + f(\mathbf{x}_2,\mathbf{v}_2)\mathbf{g}(\mathbf{x}_1)\cdot\frac{\partial \overline{f}(\mathbf{v}_1)}{\partial \mathbf{v}_1}\right\rangle. \tag{19}$$

Equation (19) is derived from equation (17) as follows. Let us take the inverse Fourier transform of equation (17):

$$\begin{aligned}
&\mathrm{i}\int \frac{\mathrm{d}^d k \mathrm{d}^d s}{(2\pi)^{2d}}\; \mathbf{s}\cdot\left[\hat{\mathbf{g}}(\mathbf{k})\hat{\overline{f}}(\mathbf{s})\hat{f}^*(\mathbf{k},\mathbf{s}) - \hat{\mathbf{g}}^*(\mathbf{k})\hat{\overline{f}}^*(\mathbf{s})\hat{f}(\mathbf{k},\mathbf{s})\right]\mathrm{e}^{-\mathrm{i}\mathbf{k}\cdot\mathbf{x}+\mathrm{i}\mathbf{s}\cdot\mathbf{v}} \\
&= \frac{\partial}{\partial \mathbf{v}}\cdot\int \frac{\mathrm{d}^d k \mathrm{d}^d s}{(2\pi)^{2d}}\; \left[\hat{\mathbf{g}}(\mathbf{k})\hat{\overline{f}}(\mathbf{s})\hat{f}^*(\mathbf{k},\mathbf{s}) - \hat{\mathbf{g}}^*(\mathbf{k})\hat{\overline{f}}^*(\mathbf{s})\hat{f}(\mathbf{k},\mathbf{s})\right]\mathrm{e}^{-\mathrm{i}\mathbf{k}\cdot\mathbf{x}+\mathrm{i}\mathbf{s}\cdot\mathbf{v}} \\
&= \frac{\partial}{\partial \mathbf{v}}\cdot\int \frac{\mathrm{d}^d k \mathrm{d}^d s}{(2\pi)^{2d}}\; \left[\hat{\mathbf{g}}(\mathbf{k})\hat{\overline{f}}(\mathbf{s})\hat{f}(-\mathbf{k},-\mathbf{s}) - \hat{\mathbf{g}}(-\mathbf{k})\hat{\overline{f}}(-\mathbf{s})\hat{f}(\mathbf{k},\mathbf{s})\right]\mathrm{e}^{-\mathrm{i}\mathbf{k}\cdot\mathbf{x}+\mathrm{i}\mathbf{s}\cdot\mathbf{v}} \\
&= 2\frac{\partial}{\partial \mathbf{v}}\cdot\int \mathrm{d}^d y \mathrm{d}^d u\; \mathbf{g}(\mathbf{y}+\mathbf{x})f(\mathbf{y},\mathbf{u})\overline{f}(\mathbf{u}+\mathbf{v}),
\end{aligned} \tag{20}$$

where the last line follows from the convolution theorem and the reality of f and $\mathbf{g}$. Fourier transforming (20) and writing $\mathbf{x}_1 = \mathbf{x}+\mathbf{y}$, $\mathbf{x}_2 = \mathbf{y}$ (and likewise for velocities) yields

$$\text{(ii)} \simeq \mathrm{i}\mathbf{s}\cdot\left[\hat{\mathbf{g}}(\mathbf{k})\hat{\overline{f}}(\mathbf{s})\hat{f}^*(\mathbf{k},\mathbf{s}) - \hat{\mathbf{g}}^*(\mathbf{k})\hat{\overline{f}}^*(\mathbf{s})\hat{f}(\mathbf{k},\mathbf{s})\right] \tag{21}$$

$$= \iint \mathrm{d}^d x_1 \mathrm{d}^d v_1 \mathrm{d}^d x_2 \mathrm{d}^d v_2 \; \mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\mathbf{x}-\mathrm{i}\mathbf{s}\cdot\mathbf{v}}\left[f(\mathbf{x}_2,\mathbf{v}_2)\mathbf{g}(\mathbf{x}_1)\cdot\frac{\partial \overline{f}(\mathbf{v}_1)}{\partial \mathbf{v}_1} + f(\mathbf{x}_1,\mathbf{v}_1)\mathbf{g}(\mathbf{x}_2)\cdot\frac{\partial \overline{f}(\mathbf{v}_2)}{\partial \mathbf{v}_2}\right], \tag{22}$$

where we have used the symmetry between 1 and 2. Ensemble averaging gives equation (19).

We will estimate $\hat{S}$ in §II E below. Note, that $\hat{S}$, as given in equations (17) and (19), need not vanish when integrated over all $\mathbf{k}$ and $\mathbf{s}$, because equation (18) is only valid in the large-k limit: non-squeezed contributions would matter at small k. Contrast this to equation (12), where there is no 'source' of C_2; effectively, equation (19) represents not an injection of total C_2 into the system, but rather a transfer of C_2 from large scales to small ones.

The three-point correlation function $\left\langle \Phi^i_j |\hat{f}|^2 \right\rangle$ is composed of Φ^i_j, which is by construction a large-scale quantity, multiplied by $|\hat{f}|^2$, which depends on $\mathbf{k}$. We contend that, as Φ^i_j varies only on large scales, it will also vary much less from one ensemble realisation to another than $|\hat{f}|^2$.

The time scales appearing on the left-hand side of equation (18) are the advection (linear) time scale associated with $\mathbf{k}\cdot(\partial/\partial\mathbf{s})$, and the non-linear (gravitational) time scale associated with $s^j\Phi^i_j\partial/\partial k^i \sim s\,\delta g_r$. We will discuss these time scales in §II C. Then, we will describe the left-hand side of (18) in §II D, and the right-hand side in §II E.

C. Critical balance

There are two (conformal) time scales in the Vlasov equation (1): linear, $\tau_\mathrm{l} \sim r/\delta v$, and non-linear, $\tau_\mathrm{nl} \sim \delta v/\delta g_r$, where δg_r is defined in equation (13) and δv is a velocity-difference scale. In the Fourier-transformed Vlasov equation (7) the linear (or phase-mixing) time scale is

$$\tau_\mathrm{l} \equiv \frac{s}{k}, \tag{23}$$

and the non-linear (gravitational) time scale is

$$\tau_\mathrm{nl} \equiv \frac{1}{s\,\delta g_r}, \tag{24}$$

where $r = 1/k$ and $s = 1/\delta v$. In the Batchelor approximation (16), the non-linear time scale is $\tau_\mathrm{nl}^{-1} = s\left\|\Phi^i_j\right\|/k$, which is the same as (24) (up to an order-unity constant).

In the highly non-linear régime, the two time scales must balance each other: if τ_l were much shorter than τ_nl, so that only the first two terms of equation (7) dominated, phase mixing would drive velocity gradients up until $\tau_{\rm nl}$ shrank to the same order of magnitude as $\tau_{\rm l}$. Conversely, if $\tau_{\rm nl}$ were much smaller, then gravitational collapse would drive spatial gradients of f, and, therefore, $\mathbf{g}$, up until $\tau_{\rm l}$ and $\tau_{\rm nl}$ matched. This so-called *critical balance* may be thought of as a type of dominant-balance asymptotic argument for the Vlasov equation, where one allows the system enough time to establish this balance (see [32, 33, 51–53] for some examples of critical balance in various areas of physics).

Critical balance amounts to setting $\tau_{\rm l} \sim \tau_{\rm nl}$. These are the only time scales in equation (18), and therefore the *cascade time* $\tau_{\rm c}$, which is the time scale for the transfer of $\hat{F}$ from scale to scale, must also be

$$\tau_{\rm c} \sim \tau_{\rm l} \sim \tau_{\rm nl}. \tag{25}$$

We will use this equality of time scales in §II F to fix the scalings of the phase-space power spectrum. Given equation (25), for each length scale $r = 1/k$, there exists a corresponding (inverse) velocity scale

$$s_{\rm c}(k) \equiv \sqrt{\frac{k}{\delta g_r}}. \tag{26}$$

It is also convenient to define $k_{\rm c}(s)$ as the inverse function of $s_{\rm c}$. If the two time scales, $\tau_{\rm l}$ and $\tau_{\rm nl}$, are equal, their shared value, $\tau_{\rm c}$, is also known as *critical-balance time*.

If $\delta g_r \sim \kappa r$, as expected for a smooth field, dominated by the outer scale, then

$$\tau_{\rm c} \sim \frac{1}{\sqrt{\kappa}}, \tag{27}$$

$$s_{\rm c}(k) = \frac{k}{\sqrt{\kappa}}, \tag{28}$$

$$k_{\rm c}(s) = \sqrt{\kappa} s. \tag{29}$$

The argument that the two time scales must balance applies only to modes well inside the horizon, such that $k \gg \mathcal{H}/c$ (where $\mathcal{H}$ is the conformal Hubble constant). Indeed, the conformal time in any asymptotically de-Sitter cosmology (i.e., one with a positive cosmological constant) is bounded from above by some value $\eta_{\rm max}$ [54]; so, for modes with k too small, linear phase mixing can only generate velocity gradients up to $s_{\rm max} \simeq k\eta_{\rm max}$. If the amplitude of $\hat{f}(k, s_{\rm max})$ is not large enough for the non-linear term to become important, then the evolution of such a mode will always be primarily linear. Here we are interested in the $k \to \infty$ limit, so it is safe to ignore this nuance. Additionally, due to the finite age of the Universe and hierarchical structure formation, the decrease of $\tau_{\rm nl}$ until it matches $\tau_{\rm l}$ might not have happened yet for all values of k, as structures on the largest scales have yet to collapse. Again, this does not affect the $k \to \infty$ limit, and we may simply take $k \gg \max\{k_{\rm nl}, \mathcal{H}/c\}$, where $k_{\rm nl}$ is defined in the introduction. In a sense, $\max\{k_{\rm nl}, \mathcal{H}/c\}$ serves as the 'outer scale' for the turbulence discussed below.

D. Phase-space cascade

Let us now see what kind of flow in $(\mathbf{k}, \mathbf{s})$ space is engendered by equation (18), ignoring $\hat{S}$ until §II E. The positive eigenvalues of $\mathrm{i}\Phi^i_j$ will drive a rotation in $(\mathbf{k}, \mathbf{s})$, where small-scale velocity structure interchanges with small-scale spatial structure, while the negative eigenvalues will drive a flow of both to ever smaller scales. In fact, one can analyse equation (16) directly—before ensemble-averaging—while still ignoring the right-hand-side. As Φ^i_j generically depends on time, this analysis holds locally in time (and space, on scales below the outer scale), but by critical balance, the long mode (large-scale) Φ^i_j cannot vary on a time scale shorter than the critical-balance time. We therefore approximate it as constant (in which case there are analytical solutions), but the qualitative features described here—namely, a rotation in $(\mathbf{k}, \mathbf{s})$ and a flow to larger values—remain also true for a time-dependent Φ^i_j. Generically, in $d > 1$, $\mathrm{i}\Phi^i_j$ would have both positive and negative eigenvalues, because this it is dominated by the large scales.

Consider a positive eigenvalue of $\mathrm{i}\Phi^i_j$. If $\mathbf{s}_+$ is its corresponding eigenvector, then for $\mathbf{s} \parallel \mathbf{s}_+$, equation (16) reduces to a transport equation under the action of a harmonic oscillator potential, i.e., a rotation in the $(\mathbf{k}, \mathbf{s})$ space. This ensures that the large-s structure in the initial condition is transported to large k, and *vice versa*. The negative eigenvalues ensure that there is a flow to ever smaller scales, because then the solution is a linear combination of hyperbolic functions (cf. [33]).

In 1D, it would appear naïvely that there is only phase-space rotation when Φ^i_j is evaluated in an initially over-dense region, because there is only one, necessarily positive, eigenvalue. This, however, is misleading. As the system evolves, matter moves around, and thus that region generically changes from being over-dense to under-dense (alternatively, it does so for different realisations of the initial conditions), and therefore the sign of $\mathrm{i}\Phi^i_j$ also changes. Thus, there is a temporal sequence of phase-space rotations and stretchings (see figure 1)—essentially, differential phase-space rotation, leading to the generation of small-scale structure (for any d).

Thus, phase-space rotations transfer small-scale structure from s to k and back, and that is supplemented by a cascade of $|\hat{f}|^2$ to ever larger k and s (similar to the plasma echo joint with a cascade in [32, 33]).[4] Hence, equation (16) describes a phase-space turbulence, where C_2, being the integral of $|\hat{f}|^2$ over scales, is cascaded to smaller scales by larger-scale tidal fields, roughly along the critical-balance line in the $(\mathbf{k}, \mathbf{s})$ space.

To test this conclusion numerically, we calculated the time-averaged phase-space power spectrum $\hat{F}$, as a func-

[4] This behaviour is generic: chaos—the exponential separation of nearby trajectories in phase-space—combined with Liouville's theorem, necessitates the formation of structure on smaller and smaller scales.

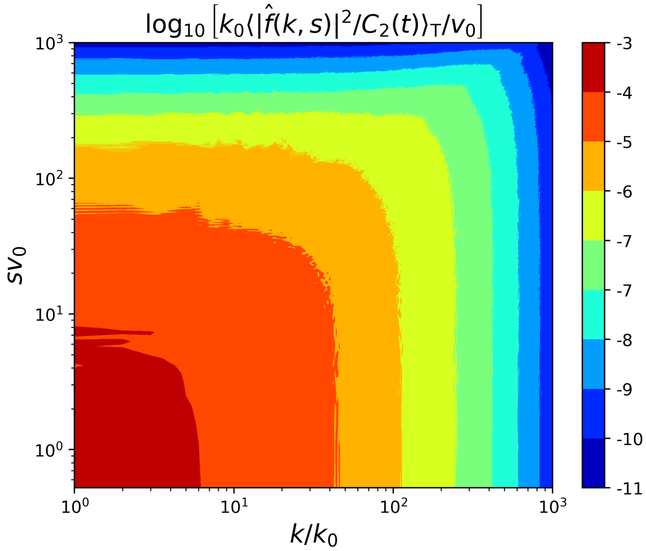

FIG. 2. A contour plot of the time-averaged power spectrum $\langle|\hat{f}|^2(t,k,s)/C_2(t)\rangle_{\rm Time}$, for the same simulation as in figure 1 (note that C_2 decays because of collisions and finite-grid effects, so it is sensible to normalise the power spectrum by C_2 at every time).

tion of $|k|$ and $|s|$ (with the negative values folded on the positive ones), for the simulation shown in figure 1. This is plotted in figure 2, which shows the contours of the time-average (as a proxy for an ensemble-average) $\left\langle|\hat{f}|^2/C_2(t)\right\rangle_{\rm Time}$. They are arranged in a rectangular shape: the critical-balance line (§II C) is the diagonal of this rectangle, and there is a slight excess along the line, manifesting the cascade. This rectangular structure is brought about by the aforementioned rotation in the (k,s) space, conjoined with the effect of the negative eigenvalues of $\mathrm{i}\Phi^i_j$, which stretch structures along the critical-balance line.

E. Source-term behaviour for a cold system

Having described the homogeneous part of equation (16), let us now consider the source term on its right-hand side.

1. Phase-space flux

Equation (18) is a conservation equation with a source, of the form

$$\frac{\partial\hat{F}}{\partial\eta}+\boldsymbol{\nabla}_{\mathbf{k},\mathbf{s}}\cdot\boldsymbol{\Gamma}=\hat{S}. \tag{30}$$

Here $\boldsymbol{\nabla}_{\mathbf{k},\mathbf{s}}$ is a 6-dimensional phase-space gradient, and $\boldsymbol{\Gamma}$ is a 6-dimensional flux, whose components are

$$\Gamma^{\mathbf{s}}_i=k_i\hat{F} \tag{31}$$

$$\Gamma^{\mathbf{k}}_i=-\mathrm{i}s^j\left\langle\Phi_{ij}|\hat{f}|^2\right\rangle. \tag{32}$$

As described in §II D, there is a flow of C_2 to larger values of k and s. Let us calculate the phase-space flux $\mathcal{F}^{\mathbf{s}}$ flowing through a sphere in $\mathbf{s}$-space of radius s, where $s_{\rm c}(k_{\rm nl})\ll s\ll v_{\rm th}^{-1}$. In steady state, integrating equation (18) over $\mathbf{k}$ yields

$$\frac{\partial}{\partial\mathbf{s}}\cdot\int\Gamma^{\mathbf{s}}\mathrm{d}^dk=\int\hat{S}\mathrm{d}^dk. \tag{33}$$

Integrating both sides over the ball in $\mathbf{s}$-space with radius s (at all $\mathbf{k}$) and using Gauss' theorem yields

$$\mathcal{F}^{\mathbf{s}}\equiv\iint\hat{s}^i\Gamma^{\mathbf{s}}_i\mathrm{d}^dk\mathrm{d}^{d-1}s=\iint_{|\mathbf{s}'|<s}\hat{S}\mathrm{d}^dk\mathrm{d}^ds', \tag{34}$$

where $\hat{s}^i=s^i/s$, and the integral d^dk is over all $\mathbf{k}$. Similarly, the flux in $\mathbf{k}$, $\mathcal{F}^{\mathbf{k}}$, is

$$\mathcal{F}^{\mathbf{k}}\equiv\iint\hat{k}^i\Gamma^{\mathbf{k}}_i\mathrm{d}^{d-1}k\mathrm{d}^ds=\iint_{|\mathbf{k}'|<k}\hat{S}\mathrm{d}^dk'\mathrm{d}^ds. \tag{35}$$

2. Jeans instability of a stream

We are now finally in a position to estimate the $(\mathbf{k},\mathbf{s})$ dependence of the flux. We will do that for cold dark matter, as described in §II A. The coldness of the distribution—its being a collection of streams—implies that each stream, in its own reference frame, is Jeans unstable, if in an over-dense region. This in turn implies that in the integral on the right-hand side of equation (35), all scales add up coherently, as we will show below.

As stated in §II A, our analysis focuses on the limit where $sv_{\rm th}\ll1\ll sv_{\rm rms}$, or, equivalently, $k_{\rm nl}\ll k\ll k_{\rm c}(v_{\rm th}^{-1})$. On these scales, f is just a collection of streams, so when zooming in on one of them, and applying the Raychaudhuri equation (5) in its frame of reference, one finds that the stream will collapse gravitationally. All the terms in equation (5) are of dimension $[\text{time}]^{-2}$, and both last terms on the right-hand-side are $O(k^2v_{\rm th}^2)$,[5] while the $\mathcal{H}\theta$ term is completely negligible, for $k\gg\mathcal{H}/c$. Equation (5) then reduces to

$$\frac{\mathrm{d}\theta}{\mathrm{d}\eta}=-\frac{\theta^2}{3}+\boldsymbol{\nabla}\cdot\mathbf{g}+O(k^2v_{\rm th}^2), \tag{36}$$

which means that for $s_{\rm c}(k)v_{\rm th}\ll1$, the gravity term $\boldsymbol{\nabla}\cdot\mathbf{g}$ dominates and there is nothing to stop a collapse. This

[5] Indeed, if initially the system had $\omega^i=0$, then, in the limit $v_{\rm th}=0$, this will stay this way locally in phase-space [41].

collapse occurs on a time scale τ_g, the *gravitational time*, defined by

$$\tau_g^{-2} \sim |\nabla \cdot \mathbf{g}| \sim 4\pi G \delta\rho_r \tag{37}$$

where again $r = 1/k$, and[6]

$$\delta\rho_r^2 \sim \int_k^\infty P(k)\mathrm{d}^d k. \tag{38}$$

This is nothing but the Jeans instability occurring on every stream individually, in its own reference frame; this is not dissimilar to the instabilities described in, e.g., [42, 45]. Indeed, a linear stability analysis of a spatially homogeneous distribution gives a Jeans-instability growth rate [37]

$$\tau_J^{-2} = \tau_g^{-2} - k^2 v_{\mathrm{th}}^2, \tag{39}$$

which is the same as τ_J^{-2} inferred from the Raychaudhuri equation (36). One special case where the collapse time may be deduced explicitly is described in appendix C.

Thus, the system experiences collapse on all spatial scales smaller than the outer scale and larger than $1/k_c(v_{\mathrm{th}}^{-1})$, so the flux (35) at a scale $k \ll k_c(v_{\mathrm{th}}^{-1})$ receives contributions from all $k' < k$; by the above discussion, structures at every scale collapse at a rate $\tau_J \sim \tau_g$, by equation (39), *provided that the system is cold on these scales*, i.e., $\tau_g k v_{\mathrm{th}} \ll 1$. In other words, C_2 from scale k' arrives at scale k at a rate $\tau_g^{-1}(k')$, so the contribution to the source from scale k'—which is $\int \hat{S}(\mathbf{k}', \mathbf{s})\mathrm{d}^d s$ by equation (35)—must scale like $\tau_g^{-1}(k')$. We thus posit that $\int \hat{S}(\mathbf{k}', \mathbf{s})\mathrm{d}^d s \sim \varepsilon/\tau_g(k')$, for a k'-independent constant ε; so for any $k \gg k_{\mathrm{nl}}$,

$$\mathcal{F}^{\mathbf{k}} = \int_0^k \mathrm{d}^d k' \int \mathrm{d}^d s\ \hat{S}(\mathbf{k}', \mathbf{s}) = \varepsilon \int_0^k \frac{\mathrm{d}^d k'}{\tau_g(k')}. \tag{40}$$

The coefficient ε has dimensions of C_2 per unit inverse volume (cf. equation (3)), and represents the amount of C_2 brought from scale k' to k. Thus, the amount of C_2 injected into scale k^{-1} by the Jeans instability, from all $k' \leq k$, is εk^d. If τ_g is proportional to a power of k, then[7]

$$\mathcal{F}^{\mathbf{k}} \sim \varepsilon \frac{k^d}{\tau_g(k)}. \tag{41}$$

By critical balance (25)

$$\tau_g \sim \tau_c \sim \tau_{\mathrm{nl}} \sim \tau_l \tag{42}$$

and, for a smooth gravitational field ($\delta g_r \sim \kappa r$), τ_g is independent of k, whence it follows that in this case $\mathcal{F}^{\mathbf{k}}$ scales as k^d. Note, however, that the argument leading to equation (41) did not require τ_g to be constant and would apply for a k-dependent τ_g.

We show in appendix B that equation (41) is *a posteriori* consistent with the scaling of the phase-space power spectrum derived from it in §II F.

F. Spectra from C_2 cascade

Now let us use our knowledge of the source, equation (41), to find the small-scale asymptotics of the phase-space power spectrum $\hat{F}$. To leading order in the large-(k, s) limit, let

$$\hat{F}(k,s) \sim \begin{cases} F_1 k^\gamma s^\xi, & \text{if } s \ll s_c(k) \ll v_{\mathrm{th}}^{-1},\ k \gg k_{\mathrm{nl}}, \\ F_2 s^\delta k^\sigma, & \text{if } k \ll k_c(s),\ s \gg v_{\mathrm{rms}}^{-1}, \end{cases} \tag{43}$$

for some $\gamma, \delta \leq 0$, $\xi, \sigma \in \mathbb{R}$. For $P(k) \propto \hat{F}(k, 0)$ to be finite, $\hat{F}$ must be independent of s in the limit $s \ll s_c$, $k_c(s) \ll k \ll k_c(v_{\mathrm{th}}^{-1})$, whence $\xi = 0$. Since $\overline{f}$ must have a defined velocity variance, $\sigma = 0$. It remains to find γ and δ.

To find γ, consider the amount of C_2 on a scale k^{-1} (or smaller), which is the variance of f over all scales up to $k = 1/r$, *viz.*,

$$\delta f_r^2 = \frac{1}{(2\pi)^{2d} V v_{\mathrm{rms}}^d} \int_{1/r}^\infty \mathrm{d}^d k \int \mathrm{d}^d s\ \hat{F}. \tag{44}$$

The flux of C_2 is given by equations (35) and (41), so the rate of change of δf_r^2 is

$$\frac{\delta f_r^2}{\tau_c} \sim \frac{\varepsilon}{V v_{\mathrm{rms}}^d} \frac{k^d}{\tau_g}, \tag{45}$$

where the cascade time τ_c is defined as the rate of change of $\hat{F}$ in equation (18).

At this point we need to invoke critical balance: we argued in §II C that, since the linear and non-linear time scales must balance each other, τ_c must also be the same as either of them—this was equation (25). It follows immediately from equation (37) that

$$\tau_c \sim \tau_g. \tag{46}$$

Therefore, from equation (45),

$$V v_{\mathrm{rms}}^d \delta f_r^2 \sim \varepsilon k^d. \tag{47}$$

The two time scales in equation (45) have cancelled, so equation (47) holds irrespectively of whether τ_g is k-dependent or not.

The s integral in (44) is dominated by $s \leq s_c(k)$ if $\hat{F}$ declines sufficiently rapidly with s (i.e., if $\delta \leq -d$, which will be verified momentarily), where $s_c(k)$ is defined in

[6] We take $k \gg \mathcal{H}/c$ and assume that $\tau_g \ll \mathcal{H}^{-1}$ so that the scale factor a in equation (2) is effectively a constant.

[7] This is indeed the case for any $0 < \lambda \leq 1$ in equation (14): by critical balance (26), one has $s_c(k) \propto k^{(1+\lambda)/2}$, so the critical-balance time is $\tau_c \propto k^{(\lambda-1)/2}$, and the collapse time has the same scaling with k.

equation (26). Inserting $\hat{F} \sim F_1 k^\gamma$ into equation (44) and then equating with (47) yields

$$\varepsilon k^d \sim F_1 s_c^d(k) k^{\gamma+d}. \tag{48}$$

This implies

$$F(k,0) \sim F_1 k^\gamma \sim \varepsilon[s_c(k)]^{-d} = \varepsilon \left(\frac{\delta g_r}{k}\right)^{d/2}. \tag{49}$$

Now let us use Poisson's equation to find γ. If the Hölder exponent in equation (14) is $\lambda < 1$, then by the Paley-Wiener theorem and equations (2) and (13), $F(k,0) \propto k^{2-d}\delta g_r^2$, whence from equation (49), $\delta g_r \propto k^{-1}$, for $d \in \{1,2,3\}$, which means that actually λ must be 1. This statement is equivalent to the statement that $\mathbf{g}$ is a smooth field, thereby justifying our choice in §II B. Inserting this into equation (49) yields $\gamma = -d$, whence follows the main result of this paper:

$$P(k) \sim \frac{\varepsilon}{V\tau_g^d} k^{-d}. \tag{50}$$

With $\lambda = 1$ now confirmed, one can find the exponent δ in a similar manner. This time, the analogue of equation (45) for the flux of C_2 to higher velocity scales is, by critical balance and equation (34),

$$\frac{\delta f_v^2}{\tau_c} \sim \frac{\varepsilon}{V v_{\rm rms}^d} \frac{[k_c(s)]^d}{\tau_g}, \tag{51}$$

where

$$\delta f_v^2 = \frac{1}{(2\pi)^{2d} V v_{\rm rms}^d} \int_{1/\delta v}^{\infty} d^d s' \int d^d k \; \hat{F} \tag{52}$$

is the variance of f over all velocity scales up to $\delta v = 1/s$. Therefore, by equation (29),

$$V v_{\rm rms}^d \delta f_v^2 \sim \varepsilon \frac{s^d}{\tau_g^d}. \tag{53}$$

If $\hat{F} \sim F_2 s^\delta$ in the limit $s \gg s_c(k)$, then, similarly to equation (48),

$$\varepsilon \frac{s^d}{\tau_g^d} \sim F_2 k_c^d(s) s^{\delta+d}, \tag{54}$$

whence $\delta = -d$.

We have thus obtained that, for cold dark matter,

$$\hat{F}(k,s) \sim \varepsilon \times \begin{cases} \tau_g^{-d} k^{-d}, & \text{if } s \ll s_c(k) \ll v_{\rm th}^{-1}, \\ s^{-d}, & \text{if } k \ll k_c(s) \ll k_c(v_{\rm th}^{-1}), \end{cases} \tag{55}$$

provided that $k \gg k_{\rm nl}$ and $s \gg v_{\rm rms}^{-1}$. These are the leading-order asymptotics: they hold for large k at $s \to 0$, and large s at $k \to 0$.[8] In general, $\hat{F}$ can also depend on the angle between $\mathbf{s}$ and $\mathbf{k}$, *viz.*, on $\mathbf{k} \cdot \mathbf{s}$. This angular dependence arises only at the next order in $s \ll s_c(k)$ or $k \ll k_c(s)$ (whichever obtains), because it must not exist at $s = 0$, or at $k = 0$.

In appendix C, we show that exactly the same scaling is obtained for an Einstein-de-Sitter background (where the scale-factor is $a(\eta) \propto \eta^2$), which has an explicit similarity symmetry, and hence an explicit way of defining the collapse time.

G. Numerical results

Let us now check whether the theoretical asymptotic scalings (55) are reproduced in our 1D Vlasov–Poisson simulation. The power spectra of the system illustrated in figure 1 are shown in figure 3: a k^{-1} power law establishes itself quickly, persisting up to $k \sim 100k_0$, which is of the order of $k_c(v_{\rm th}^{-1})$, as expected. These spectra are just $|\hat{f}|^2/C_2(t)$, averaged over short time intervals (compared with the simulation's duration) as a proxy for the ensemble average. We re-scale $|\hat{f}|^2$ by $C_2(t)$ at every time step, because C_2 decays due to collisions and finite-grid effects. In figure 4, we average over the entire simulation, displaying cleaner power laws in the appropriate range of k and s. While the system is not stationary, this only affects the amplitude—not the overall scaling—and figure 3 shows that the re-scaling by C_2 corrects for that; hence, we can trust figure 4 to give an adequate approximation for $\hat{F}$.

As the system is sourced by the gravitational-collapse instability, which was shown in §II E to continue until $(kv_{\rm th})^{-1}$ matches the collapse time, the scaling of the power spectrum must be truncated at $v_{\rm th}$. To test the theory further, it is necessary to see whether it is indeed the case that the asymptotics (55) persist until $v_{\rm th}$ is reached. We ran identical simulations, differing only by the value of $v_{\rm th}$, to verify this. The result is presented in figure 5: the left column has $v_{\rm th} = 0.01v_0$ while the right column has $v_{\rm th} = 0.05v_0$. The s-spectra in the bottom row show that indeed, the s^{-1} power law is truncated at a lower value of s, by a factor that matches the ratio of $v_{\rm th}$ for the two runs. The s^{-2} (or k^{-2}) scaling of $\hat{F}$ at $sv_{\rm th} \gg 1$ (or $k \gg k_c(v_{\rm th}^{-1})$) is a known result in plasma systems [33], discussed in appendix A 3.

The k^{-d} asymptotic of the density power spectrum may be derived by an altogether different, yet systematic approach—by performing an asymptotic analysis of the an integral expression for $P(k)$ and examining its critical points. We do so in §III below, which is self-contained.

III. SADDLE-POINT APPROACH

In this section, we focus on $d = 3$ spatial dimensions. For pure gravitational Newtonian evolution of N identi-

[8] This scaling is marginal, in that the contribution of the region up to $s = s_c$ in the integral in equation (44) turns out to be formally as large in the region $s > s_c$; this, however, does not invalidate the conclusion, because if both are dominant, then they both contribute $\sim s_c^d$, so the rest of the argument still goes through.

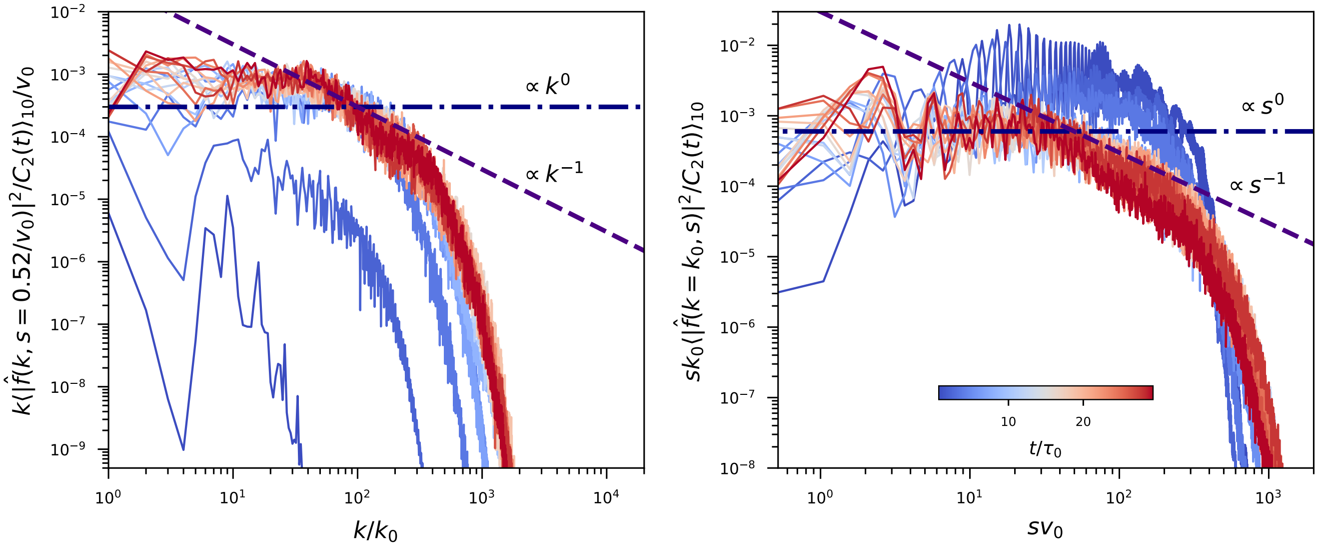

FIG. 3. The time evolution of the power spectra $\langle|\hat{f}|^2/C_2(t)\rangle_{10}$ for the simulation illustrated in figure 1 (early times: blue; late times: red). The time averaging was over 10 simulation outputs (corresponding to time windows of duration $1.36\tau_0$). *Left panel*: the k-spectrum at $s = 0.52/v_0$. *Right panel*: the s-spectrum at $k = k_0$. Both spectra are compensated by the expected asymptotics (multiplied by k and s, respectively). A k^{-1} power law establishes itself quickly, extending to $k \sim 200k_0$, which is of the order of $k_\mathrm{c}(v_\mathrm{th}^{-1})$, as expected; similarly, the s^{-1} power law extends to $sv_0 \sim 200$.

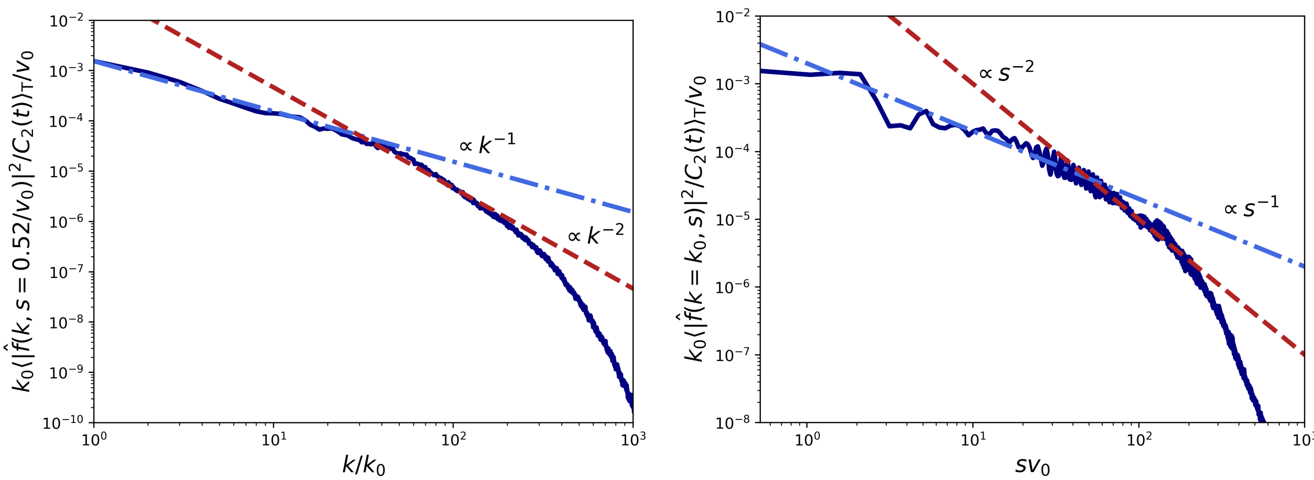

FIG. 4. The power spectra for the same simulation as in figure 1, but now averaged over $t \in [0, 30]\tau_0$, as a proxy for the ensemble average. Unlike in figure 3, they are uncompensated, and display the theoretical k^{-d} and s^{-d} scalings explicitly ($d = 1$).

cal particles, the power spectrum is given exactly by [34]

$$P(k,t) = \frac{M^2}{V}\prod_{n=1}^{N}\int \mathrm{d}^3q_n\mathrm{d}^3p_n\mathcal{P}(\{q\}, \{p\})\mathrm{e}^{\mathrm{i}\mathbf{k}\cdot[\mathbf{x}_1(t)-\mathbf{x}_2(t)]}, \tag{56}$$

where $(\mathbf{q}_n, \mathbf{p}_n)$ is the initial phase-space position of particle n, $(\mathbf{x}_n(t), \mathbf{v}_n(t))$ is the phase-space position of particle n at time t, and $\mathcal{P}(\{q\}, \{p\})$ is the joint probability distribution of the initial phase-space positions $(\{q\}, \{p\}) \equiv \{(\mathbf{p}_n, \mathbf{q}_n)\}_{n=1}^N$ of all particles. This equation is permutation-invariant, and, therefore, the choice of two particles is arbitrary.

For *cold* dark matter with Gaussian initial conditions, the initial distribution $\mathcal{P}$ is

$$\mathcal{P}(\{q\}, \{p\}) = \frac{V^{-N}\mathcal{C}(\{q\}, \{p\})}{\sqrt{(2\pi)^{3N}\det C_{pp}^N}}\mathrm{e}^{-\{p\}^T(C_{pp}^N)^{-1}\{p\}/2}, \tag{57}$$

where $C_{pp}^N = C_{pp}^N(\{q\})$ is the $3N \times 3N$ covariance matrix of $\{p\}$, and $\mathcal{C}$ encapsulates initial density-density and density-momentum correlations [55], both of whose functional dependence on particle positions depends on the

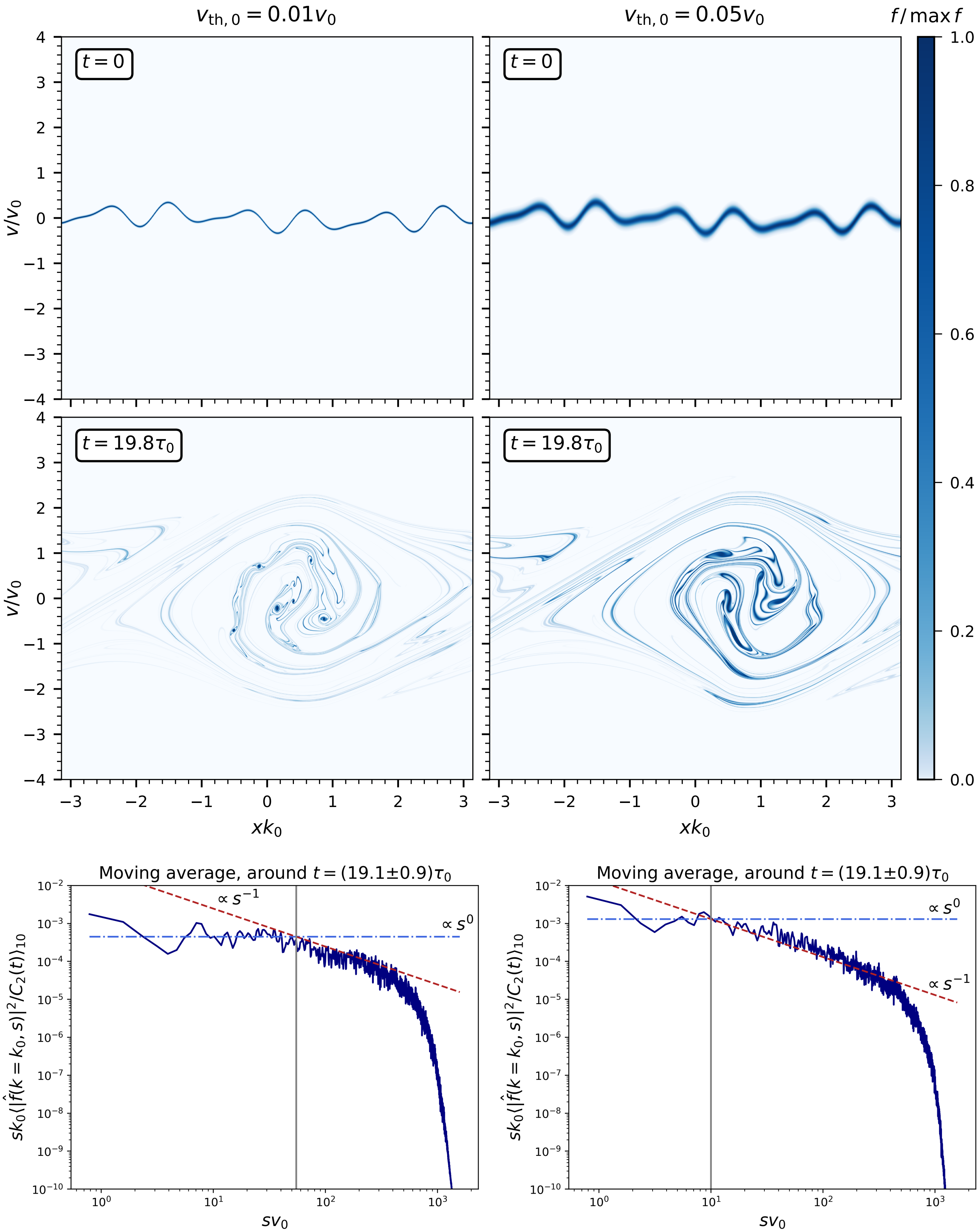

FIG. 5. A comparison between the evolution and the spectra of two systems with identical initial conditions, save for the value of v_{th}. *Left panels*: a cold case, $v_{\mathrm{th}} = 0.01v_0$; *right panels*: a warmer case, $v_{\mathrm{th}} = 0.05v_0$. The plots in the bottom row show the s spectra at the end of the simulations, compensated by s. In the warmer case, the spectrum ceases to scale as s^{-1} at around $sv_0 \sim 10$, while in the colder one, it does so at $sv_0 \sim 50$ (marked by grey lines). In both cases the s^{-1} scaling is replaced at $sv_{\mathrm{th}} \gg 1$ by s^{-2}. See text and appendix A for details. Videos are available here.

cosmology (we take a ΛCDM background).

The usual procedure to obtain the power spectrum $P(k)$ from equation (56) would involve integrating out particles $3, \ldots, N$, leaving only a 12-dimensional integral, over the phase-sub-space of particles 1 and 2. We will, however, go the other way round, and integrate *first* over the relative position and momentum of this pair, and only then over all other particles—we will see that this order of integration is well suited to deriving the asymptotics of $P(k,t)$ as $k \to \infty$. We change the integration variables from $\mathbf{q}_1$ and $\mathbf{q}_2$ to $\mathbf{q} \equiv \mathbf{q}_1 - \mathbf{q}_2$ and $\mathbf{Q} \equiv (\mathbf{q}_1 + \mathbf{q}_2)/2$, and to their conjugate momenta, $\mathbf{p}$ and $\mathbf{P}$, respectively. As $\{p\}^T \,(C^N_{pp})^{-1}\,\{p\}$ is quadratic in $\{p\}$, we can make the $\mathbf{p}$ dependence explicit, *viz.*,

$$\begin{aligned}&\{p\}^T \,(C^N_{pp})^{-1}\,\{p\} \equiv \mathbf{p}^T \Sigma^{-1}(\mathbf{q})\,\mathbf{p} \\ &- 2\mathbf{a}\left(\mathbf{q},\mathbf{Q},\mathbf{P},\{(q,p)\}_3^N\right)\cdot\mathbf{p} - 2B\left(\mathbf{q},\mathbf{Q},\mathbf{P},\{(q,p)\}_3^N\right),\end{aligned} \tag{58}$$

where $\mathbf{a}$ is a linear function of momenta and B a quadratic one. Below we will require the following properties of C^N_{pp}, derived in appendix D, based on the assumption $v_{\rm th} \to 0$ (so the conclusions again will be valid at $k \ll k_{\rm c}(v_{\rm th}^{-1})$, as in §II B 1): (i) $\Sigma \simeq Aq^2$ at small q for some order-unity matrix A, (ii) $|\mathbf{a}| \sim O(q^{-1})$ (at most) in this limit, and (iii) at $q \to \infty$, the correlation matrix C^N_{pp} tends to a constant, i.e., Σ is order unity in the limit $q \to \infty$.

Using Duhamel's principle for Hamilton's equations of motion, one has [e.g., 34]

$$\mathrm{i}\mathbf{k}\cdot[\mathbf{x}_1(t) - \mathbf{x}_2(t)] = \mathrm{i}\mathbf{q}\cdot\mathbf{k} + \mathrm{i}g_{qp}\left(t, t_{\rm initial}\right)\mathbf{p}\cdot\mathbf{k} + \mathrm{i}k\psi_I, \tag{59}$$

where g_{qp} is a 'propagator' that encapsulates the linear evolution (defined below), and the 'interaction term' is

$$\begin{aligned}&k\psi_I\left(\mathbf{P},\mathbf{Q},\{(q,p)\}_3^N\right) \\ &\equiv \mathbf{k}\cdot\int_0^t \mathrm{d}t' g_{qp}(t', t_{\rm initial})\left(\mathbf{F}_1(t') - \mathbf{F}_2(t')\right),\end{aligned} \tag{60}$$

$\mathbf{F}_n$ being the additional acceleration of particle n, relative to its motion in the part already included in g_{qp}, caused by all other particles. For example, if $g_{qp} = t - t_{\rm initial}$, one re-obtains regular Newtonian dynamics. Using results of first-order cosmological perturbation theory (or, alternatively, re-summed kinetic field theory [56]), we take $g_{qp}(t) \equiv [D_+(t) - D_+(t_{\rm initial})]\,/\,[\mathrm{d}D_+/\mathrm{d}t]$, where D_+ is the ΛCDM growth factor. Then ignoring the ψ_I term in (59) would just yield the Zel'dovich approximation, whose asymptotics were studied by [57, 58].

A. Possible saddle points

Using equations (58) and (59), we define the exponent

$$\varphi \equiv -\frac{1}{2}\mathbf{p}^T\Sigma^{-1}\mathbf{p} + \mathbf{a}\cdot\mathbf{p} + \mathrm{i}\mathbf{q}\cdot\mathbf{k} + \mathrm{i}g_{qp}\mathbf{p}\cdot\mathbf{k} + \mathrm{i}k\psi_I + B, \tag{61}$$

so the integral (56) for the power spectrum turns into

$$P(k,t) \propto \prod_{n=3}^{N}\int \frac{\mathrm{d}^3q_n\mathrm{d}^3p_n\mathrm{d}^3Q\mathrm{d}^3P\mathrm{d}^3q\mathrm{d}^3p}{V^N\sqrt{(2\pi)^{3N}\det C^N_{pp}}}\mathcal{C}(\{q\},\{p\})\mathrm{e}^{\varphi}. \tag{62}$$

We start by integrating over $\mathbf{p}$ and $\mathbf{q}$, applying the saddle-point approximation. This is a movable-saddle problem, so it requires care in handling [59, 60].[9] The function ψ_I is a smooth function of $\mathbf{q}$ and $\mathbf{p}$, because it arises from the Hamiltonian flow in phase-space, generated by a smooth gravitational potential (recall that we neglect collisions and that time is bounded). This assertion is just the statement that $\lambda = 1$ as in §II B 1. Letting $\boldsymbol{\psi}_q \equiv \partial\psi_I/\partial\mathbf{q}$ and $\boldsymbol{\psi}_p \equiv \partial\psi_I/\partial\mathbf{p}$, the exponent (61) is stationary when

$$\begin{aligned}&-\Sigma^{-1}\mathbf{p} + \mathbf{a} + \mathrm{i}g_{qp}\mathbf{k} + \mathrm{i}k\boldsymbol{\psi}_p = 0, \\ &-\frac{1}{2}p^i\frac{\partial(\Sigma^{-1})_{ij}}{\partial\mathbf{q}}p^j + p_i\frac{\partial a^i}{\partial\mathbf{q}} + \frac{\partial B}{\partial\mathbf{q}} + \mathrm{i}\mathbf{k} + \mathrm{i}k\boldsymbol{\psi}_q = 0.\end{aligned} \tag{63}$$

Let us parameterise the solution of these equations as

$$\mathbf{p} = k^{\alpha}\mathbf{c}, \;\; \mathbf{q} = k^{\beta}\mathbf{d}, \tag{64}$$

where $\mathbf{c}$ and $\mathbf{d} \in \mathbb{C}^3$ are complex vectors whose magnitude remains finite as $k/k_{\rm nl} \to \infty$. Together with α, β, they are to be determined by equations (63), by seeking a dominant balance, i.e., such a balance that the exponent (61) has the weakest k dependence around the stationary point that solves equations (63).

A priori, equations (63) could permit a balance that is independent of the initial-condition distribution, i.e., that same balance would exist for uniform initial conditions. But in that case, Liouville's theorem—after changing variables from the initial phase-space positions $\{(\mathbf{q}_n,\mathbf{p}_n)\}^N$ to the current ones $\{(\mathbf{x}_n,\mathbf{v}_n)\}^N$—ensures that this yields a contribution to $P(k)$ proportional to $\delta^{\rm D}(\mathbf{k})$, whence this balance is not dominant. Conversely, the dominant balance must involve the first terms on the left-hand sides of equations (63). If $\beta > 0$, then, as $\Sigma \underset{q\to\infty}{\longrightarrow}$ const, the first term in the first of equations (63) is proportional to p. As this must play a part in a dominant balance, this implies that $\alpha = 1$. Substituting such a saddle point into equation (61) gives an exponentially suppressed contribution, $\sim O\left[\exp(\propto -k^2)\right]$ to $P(k)$ at most.

Thus, stationary points with $\beta < 0$ are dominant, as they could contribute a power-law tail to $P(k)$. For $\beta < 0$, equations (63) imply $1 = \alpha - 2\beta$ and $1 = 2\alpha - 3\beta$, whence $\alpha = \beta = -1$. This is consistent as long as

$$\lim_{\substack{(q,p)\sim k^{-1}\\ k\to\infty}}\left(\left|\boldsymbol{\psi}_q\right|, \left|\boldsymbol{\psi}_p\right|\right) = O(1), \tag{65}$$

[9] Note also that the saddle point may be complex, but that is innocuous, since the exponent can be continued analytically to the complex plane.

which is generically valid, and so the interaction term is potentially as important as the linear term; that both are equally important is essentially a statement of critical balance (cf. §II C).

B. Evaluation of the asymptotics

Having proven that the asymptotic expansion of the power spectrum is given, up to exponentially small (in k) contributions, by its dominant saddle point at $q, p \sim k^{-1}$, one is allowed to replace φ with its expansion at small q and p. This yields

$$\varphi \simeq -\frac{1}{2}\mathbf{p}^T\Sigma^{-1}\mathbf{p} + \mathbf{a}\cdot\mathbf{p} + \mathrm{i}\mathbf{q}\cdot\mathbf{k} + \mathrm{i}g_{qp}\mathbf{p}\cdot\mathbf{k} + \mathrm{i}k\left[\mathbf{p}\cdot\boldsymbol{\psi}_p(0) + \mathbf{q}\cdot\boldsymbol{\psi}_q(0)\right] + B(q=0) + o(1). \tag{66}$$

With this asymptotic approximation for φ, we need to integrate equation (62) over $\mathbf{p}$ and $\mathbf{q}$ to obtain its asymptotic scaling with k. Observe that $\det C_{pp}^N$ factorises into $\det\Sigma$ multiplied by normalisation factors for the other momenta variables. The only other pieces that still depend on $\mathbf{Q}$, $\mathbf{P}$, and the initial positions and momenta of the other particles are $\mathcal{C}$, $\mathbf{a}$, B, and $\boldsymbol{\psi}_{q,p}$. Thus,

$$P(k,t) \simeq \frac{M^2}{V^2}\left\langle\int \frac{\mathrm{d}^3q\mathrm{d}^3p\ \mathcal{C}}{\left[(2\pi)^3\det\Sigma\right]^{1/2}}\exp\left(-\frac{1}{2}\mathbf{p}^T\Sigma^{-1}\mathbf{p} + \mathbf{a}\cdot\mathbf{p} + \mathrm{i}\mathbf{q}\cdot\mathbf{k} + \mathrm{i}g_{qp}\mathbf{p}\cdot\mathbf{k}\right)e^{\mathrm{i}k\left[\mathbf{p}\cdot\boldsymbol{\psi}_p(0) + \mathbf{q}\cdot\boldsymbol{\psi}_q(0)\right]}\right\rangle, \tag{67}$$

where the average is over the position and momentum of the centre of mass of particles 1 and 2, as well as the positions and momenta of all the other particles (e^B is absorbed into the average). Integrating over $\mathbf{p}$, we get

$$P(k,t) \simeq \frac{M^2}{V^2}\left\langle\int \mathrm{d}^3q\ \mathcal{C}\ \exp\left\{-\frac{k^2}{2}\left[g_{qp}\hat{\boldsymbol{k}} + \boldsymbol{\psi}_p(0) - \frac{\mathrm{i}\mathbf{a}}{k}\right]^T\Sigma(\mathbf{q})\left[g_{qp}\hat{\boldsymbol{k}} + \boldsymbol{\psi}_p(0) - \frac{\mathrm{i}\mathbf{a}}{k}\right] + \mathrm{i}\mathbf{q}\cdot\mathbf{k} + \mathrm{i}k\mathbf{q}\cdot\boldsymbol{\psi}_q(0)\right\}\right\rangle. \tag{68}$$

This integration does not introduce any powers of k because both d^3p and $\sqrt{\det\Sigma}$ are proportional to k^{-3} in the vicinity of the stationary point, so the two cancel. Recalling that the saddle point is at $\mathbf{q} = \mathbf{d}/k$, where $\mathbf{d}$ is a finite constant, we change variables to $\mathbf{y} = k\mathbf{q}$, whence

$$P(k,t) \simeq \frac{M^2}{V^2k^3}\left\langle\int \mathrm{d}^3y\ \mathcal{C}\ \mathrm{e}^{-k^2\left[g_{qp}\hat{\boldsymbol{k}} + \boldsymbol{\psi}_p(0) - \frac{\mathrm{i}}{k}\mathbf{a}(\mathbf{y}/k)\right]^T\Sigma(\mathbf{y}/k)\left[g_{qp}\hat{\boldsymbol{k}} + \boldsymbol{\psi}_p(0) - \frac{\mathrm{i}}{k}\mathbf{a}(\mathbf{y}/k)\right] + \mathrm{i}\mathbf{y}\cdot\left[\hat{\boldsymbol{k}} + \boldsymbol{\psi}_q(0)\right]/2}\right\rangle \equiv \frac{F_1}{V}k^{-3}. \tag{69}$$

Therefore, the integral d^3q, as opposed to the momentum integral in (67), is not compensated by any function that scales like k^{-3}, so it is this integration that yields the k^{-3} scaling, which emerges from $q \sim k^{-1}$ at the stationary point. The exponent in equation (69) is order unity in the limit $k \to \infty$, because $\Sigma \sim y^2/k^2$, and also because $\mathbf{a}/k$ is at most order unity,[10] whence the y integral is a Gaussian integral, and the entire factor multiplying k^{-3} in equation (69), denoted by F_1/V, is of order unity.

Thus, we have established that for collisionless, cold dark matter in the non-relativistic limit, the asymptotic behaviour of the power spectrum as $k \to \infty$ is dominated by the contribution of the saddle point at $q, p \sim k^{-1}$. This yields $P(k) \sim k^{-3}$, with an order-unity coefficient; that this coefficient is non-zero follows from the qualitative argument of §II.

[10] The exponent might depend on $\hat{\boldsymbol{k}}$, the direction of $\mathbf{k}$, before being ensemble-averaged.

IV. DISCUSSION

A. The importance of being cold

The k^{-d} power spectrum arose from a turbulent cascade in phase-space, where the gravitational field was treated as a smooth field—this is an analogue of Batchelor turbulence in hydrodynamics [49], where a tracer is advected by a large-scale, smooth velocity field. A crucial step in the derivation of the source term in §II E was that the source was $\int \mathrm{d}^d s\ \hat{S} \sim \text{const}$, at $v_{\text{rms}}^{-1} \ll s_{\text{c}}(k), s \ll v_{\text{th}}^{-1}$, i.e., that it accumulated when integrated over $\mathbf{k}$, to give $\varepsilon k^d/\tau_{\text{g}}$. This was also necessary in §III, where the initial conditions were special to cold dark matter, with the particular functional form of the C_{pp}^N matrix, $\Sigma \sim Aq^2$, reflecting the fact that if two particles start out at exactly the same spatial position, they will remain together forever.

When the system is not cold, these assumptions fail. Indeed, the source term would not accumulate when integrated over k and s up to small, non-linear scales, because at $sv_{\rm th} \gg 1$ and $s_c(k)v_{\rm th} \gg 1$, there is no instability to drive the collapse, so one would expect $\iint \hat{S}\mathrm{d}k\mathrm{d}s = \mathrm{const.}$

The coldness of the system has two implications: first, it means that locally in phase-space, the distribution function is always a stream, and therefore, locally, the Raychaudhuri equation (5) applies. This, in turn, means that the stream is unstable to gravitational collapse, with the collapse time scale $\tau_{\rm g} \sim \tau_{\rm J}$, precisely the time scale that enters in equation (45). Secondly, this time scale is scale-independent, so the contributions to the integrated source (34) add up constructively. Together, these two features imply that the source has the form (41); this then enables the estimate of $\hat{F}$ in §II F.

While in a collisionless setting, $v_{\rm th}$ would not change with time,[11] in reality finite-N effects do increase $v_{\rm th}$. Additionally, baryons would influence the matter power spectrum significantly on small scales, probably leading to a deviation from the power-law scalings found here. The approach used in this paper does not apply once interactions with baryons are included, and this is its main limitation. Another limitation is that it only applies in-so-far as dark matter behaves as a *cold*, classical *phase-space* fluid—other dark-matter candidates might follow different scalings, and we defer this to future work.

Were it not for baryons, a full theory of the constituent particle of dark matter should predict both the initial value of $v_{\rm th}$ and the functional form of the effective collision operator; this should allow one to find how C_2 changes with time, and thence how $v_{\rm th}$ evolves. Measuring the break in the dark-matter power spectrum—the transition from k^{-3} to a different power law (in $d = 1$, to a k^{-2} scaling, as in figure 4)—would hypothetically allow one to find the present-day value of $v_{\rm th}$, which is directly connected to the nature of dark matter (if indeed it held that $u_\nu \ll v_{\rm th}$, where the collisional velocity scale u_ν is defined in appendix A). For example, for WIMPs, which decouple from the photons when non-relativistic, one would have

$$v_{\rm th}(z) = c\frac{T(z)}{T_{\rm kd}}\sqrt{\frac{2k_{\rm B}T_{kd}}{mc^2}} \approx 3.3\times 10^{-12}c(1+z)\left[\frac{10\ \mathrm{MeV}}{T_{\rm kd}}\right]^{1/2}\left[\frac{\mathrm{GeV}}{m}\right]^{1/2}, \tag{70}$$

where $T(z)$ is the photon temperature at redshift z and $T_{\rm kd}$ is the kinetic-decoupling temperature [21, 22, 61]. This is far too small to be observed practically, and moreover, it corresponds to (cf. appendix C)

$$k_{\rm c}(v_{\rm th}^{-1}) \sim \delta_{\rm typ}^{1/2}\frac{H}{v_{\rm th}} \approx \frac{10^8 h(z)}{1+z}\delta_{\rm typ}^{1/2}\ \mathrm{Mpc}^{-1}\left[\frac{m}{\mathrm{GeV}}\right]^{1/2}\left[\frac{T_{\rm kd}}{10\ \mathrm{MeV}}\right]^{1/2}, \tag{71}$$

where $\delta_{\rm typ}$ is a typical over-density and the Hubble constant is $H(z) = 100h(z)$ km s^{-1} Mpc^{-1} (this value is of the same order as the free-streaming scale [21] for $\delta_{\rm typ} \sim 1$).

B. Some like it hot: similarities with plasma physics

A phase-space Batchelor-type cascade was recently proposed to be the universal régime of (plasma) Vlasov–Poisson turbulence at Debye and sub-Debye scales [32, 33]. This has been verified numerically in 1D simulations of turbulence driven by external forcing [33] and the two-stream instability [62]. Like in the cold-dark-matter turbulence presented here, the (electric) field fluctuations are spatially smooth, so the phase-space mixing of the distribution function is dominated by the outer-scale fields. However, unlike cold-dark-matter turbulence, which is sourced at every Jeans-unstable scale, the cascade in the plasma case has a constant flux of C_2, because C_2 is only sourced at the outer scale; this changes the scalings of the power spectrum and of the field spectrum. This situation is analogous to the gravitational phase-space turbulence at $s \gg v_{\rm th}^{-1}$ that appears in the simulations presented in figure 5. We will investigate this régime further in future work.

C. Implications for dark-matter haloes

The $P(k) \propto k^{-3}$ scaling (in 3D) derived here for the non-linear power spectrum sheds some light on universal properties of dark matter haloes. Within the hierarchical-clustering paradigm [63], the small-scale limit of the power spectrum is dominated by the one-halo term,

$$P_{1h}(k) = \int \mathrm{d}M\ \frac{\mathrm{d}n}{\mathrm{d}M}[R^3\overline{\delta}\hat{u}(kR)]^2, \tag{72}$$

where $\mathrm{d}n/\mathrm{d}M$ is the halo mass function, $R(M) \equiv R_{200}(M)/c(M)$, $\overline{\delta}$ is the amplitude of the halo density profile, $c(M)$ is the halo's concentration, R_{200} is the radius where the density is 200 times the critical density of the Universe, and $\hat{u}$ is the Fourier transform of the normalised density profile $u(x/R)$ [64]. If we define $\nabla_c \equiv -\mathrm{d}\ln c/\mathrm{d}\ln M$ and $\nabla_n \equiv \mathrm{d}\ln(\mathrm{d}n/\mathrm{d}M)/\mathrm{d}\ln M$, both at $M \to 0$, then, as $k \to \infty$

$$P_{1h}(k) \sim k^\gamma, \qquad \gamma = -3\frac{3+\nabla_n}{1+3\nabla_c}, \tag{73}$$

[11] Because $C_2 \sim \|f\|_\infty^2 v_{\rm th}^d V$, and both C_2 and $\|f\|_\infty$ are conserved (in fact, one can define $v_{\rm th} \equiv C_2^{1/d}\|f\|_\infty^{-2/d}V^{-1/d}/\sqrt{\pi}$, as a measure of the thinness of the distribution). There is of course a change in $\|f\|_\infty = \max f$ because of the Universe's expansion, but this is very slow for the large-k limit, where $k \gg \mathcal{H}/c$.

provided the halo density profile $\hat{u}$ decreases sufficiently fast with its argument. Setting $\gamma = -3$, as would be required by our result, yields a relation between the universal concentration-to-mass relation and the halo mass function for cold dark matter:

$$\nabla_n = 3\nabla_c - 2; \tag{74}$$

this constrains semi-analytic prescriptions for halo mass-functions [65].

V. SUMMARY

In this paper we have described a physical mechanism that produces the k^{-d} asymptotic scaling of the dark-matter power spectrum naturally. This was done in two ways: by expressing $P(k)$ as a phase-space integral and analysing it with a stationary-phase method (§III), and via a phenomenological study of a critically balanced phase-space cascade, akin to Batchelor turbulence, for a cold, collisionless, self-gravitating system (§II). Both methods are phase-space based, and so remain valid even after streams cross. The fact that the phase-space distribution function is cold (i.e. that it is only non-zero in a d-dimensional sub-manifold of phase-space) was crucial to both approaches. Gravitational collapse sources a cascade of the quadratic Casimir invariant. The cascade is sustained by the joint action of phase mixing and tidal forces (by smooth fields), which together transfer phase-space structure into ever smaller scales. The balance between linear free streaming and tidal forces also appears in the saddle-point argument of §III. Usually in turbulent systems there exists an inertial range of scales [66] where there is a constant-flux cascade of an invariant. Gravitational turbulence is different in that the flux is not constant over the range of scales of interest, and yet there is a universal scaling régime of the phase-space power spectrum.

The validity of our approach is supported by 1D Vlasov–Poisson simulations, which confirm our theoretical predictions.

Our determination of the small-scale asymptotics of the dark-matter power spectrum may allow for a non-trivial test of effective field theories of the large scales, either by imposing these asymptotics on them, or by using $\hat{F}$ found here as a closure for these theories. We intend to investigate these possibilities in future work.

ACKNOWLEDGMENTS

We wish to thank Michael Barnes, William Clarke, and Anatoly Spitkovsky for helpful discussions, and are grateful to James Juno for help and advice with the Gkeyll code. Y.B.G. was supported in part by a Leverhulme Trust International Professorship Grant to S. L. Sondhi (No. LIP-2020-014). Y.B.G. and A.A.S. were supported in part by a Simons Investigator Award to A.A.S. from the Simons Foundation. M.L.N. was supported by a Clarendon Scholarship at Oxford. R.J.E. was supported by a UK EPSRC studentship. S.K. and M.B. were supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy EXC 2181/1–390900948 (the Heidelberg STRUCTURES Excellence Cluster). This research was also supported in part by NSF grant PHY-2309135 to the Kavli Institute for Theoretical Physics (KITP).

[1] P. J. E. Peebles, *The Large-Scale Structure of the Universe* (Princeton University Press, Princeton, N.J., 1980).
[2] S. Dodelson, *Modern Cosmology* (Academic Press, Amsterdam, 2003).
[3] V. Desjacques, D. Jeong, and F. Schmidt, Large-scale galaxy bias, Phys. Rep. **733**, 1 (2018), arXiv:1611.09787 [astro-ph.CO].
[4] N. Aghanim, Y. Akrami, M. Ashdown, J. Aumont, C. Baccigalupi, M. Ballardini, A. J. Banday, R. B. Barreiro, N. Bartolo, S. Basak, *et al.* (Planck Collaboration), Planck 2018 results. VI. Cosmological parameters, A&A **641**, A6 (2020), arXiv:1807.06209 [astro-ph.CO].
[5] M. Tegmark and M. Zaldarriaga, Separating the early universe from the late universe: Cosmological parameter estimation beyond the black box, Phys. Rev. D **66**, 103508 (2002), arXiv:astro-ph/0207047 [astro-ph].
[6] G. Cabass, M. M. Ivanov, M. Lewandowski, M. Mirbabayi, and M. Simonović, Snowmass white paper: Effective field theories in cosmology, Physics of the Dark Universe **40**, 101193 (2023), arXiv:2203.08232 [astro-ph.CO].
[7] D. Heggie and P. Hut, *The Gravitational Million–Body Problem: A Multidisciplinary Approach to Star Cluster Dynamics* (Cambridge University Press, Cambridge, 2003).
[8] A. J. S. Hamilton, P. Kumar, E. Lu, and A. Matthews, Reconstructing the primordial spectrum of fluctuations of the Universe from the observed nonlinear clustering of galaxies, ApJ **374**, L1 (1991).
[9] Y. B. Ginat, Multiple-scales approach to the averaging problem in cosmology, J. Cosmology Astropart. Phys. **2021**, 049 (2021), arXiv:2005.03026 [gr-qc].
[10] D. Baumann, A. Nicolis, L. Senatore, and M. Zaldarriaga, Cosmological non-linearities as an effective fluid, J. Cosmology Astropart. Phys. **2012**, 051 (2012), arXiv:1004.2488 [astro-ph.CO].
[11] T. Baldauf, Effective field theory of large-scale structure, in *Effective Field Theory in Particle Physics and Cosmology: Lecture Notes of the Les Houches Summer School: Volume 108, July 2017* (Oxford University Press, 2020).
[12] M. M. Ivanov, Effective field theory for large-scale

structure, in *Handbook of Quantum Gravity*, edited by C. Bambi, L. Modesto, and I. Shapiro (Springer Nature Singapore, Singapore, 2023).
[13] V. Springel, R. Pakmor, A. Pillepich, R. Weinberger, D. Nelson, L. Hernquist, M. Vogelsberger, S. Genel, P. Torrey, F. Marinacci, and J. Naiman, First results from the IllustrisTNG simulations: matter and galaxy clustering, MNRAS **475**, 676 (2018), arXiv:1707.03397 [astro-ph.GA].
[14] A. Taruya and S. Colombi, Post-collapse perturbation theory in 1D cosmology—beyond shell-crossing, MNRAS **470**, 4858 (2017), arXiv:1701.09088 [astro-ph.CO].
[15] A. Halle, S. Colombi, and S. Peirani, Phase-space structure analysis of self-gravitating collisionless spherical systems, A&A **621**, A8 (2019), arXiv:1701.01384 [astro-ph.GA].
[16] S.-F. Chen and M. Pietroni, Asymptotic expansions for Large Scale Structure, J. Cosmology Astropart. Phys. **2020**, 033 (2020), arXiv:2002.11357 [astro-ph.CO].
[17] I. H. Gilbert, Collisional relaxation in stellar systems, ApJ **152**, 1043 (1968).
[18] P.-H. Chavanis, Kinetic theory of long-range interacting systems with angle-action variables and collective effects, Physica A **391**, 3680 (2012), arXiv:1107.1475 [cond-mat.stat-mech].
[19] D. Lazarovici and P. Pickl, A mean field limit for the Vlasov-Poisson system, Arch. Ration. Mech. Anal. **225**, 1201 (2017).
[20] C. Rampf, Cosmological Vlasov-Poisson equations for dark matter, Rev. Mod. Plasma Phys. **5**, 10 (2021).
[21] A. M. Green, S. Hofmann, and D. J. Schwarz, The power spectrum of SUSY-CDM on subgalactic scales, MNRAS **353**, L23 (2004), arXiv:astro-ph/0309621 [astro-ph].
[22] A. Loeb and M. Zaldarriaga, Small-scale power spectrum of cold dark matter, Phys. Rev. D **71**, 103520 (2005), arXiv:astro-ph/0504112 [astro-ph].
[23] M. Kunz, S. Nesseris, and I. Sawicki, Constraints on dark-matter properties from large-scale structure, Phys. Rev. D **94**, 023510 (2016), arXiv:1604.05701 [astro-ph.CO].
[24] D. B. Thomas, M. Kopp, and C. Skordis, Constraining the properties of dark matter with observations of the Cosmic Microwave Background, ApJ **830**, 155 (2016), arXiv:1601.05097 [astro-ph.CO].
[25] D. Gilman, S. Birrer, A. Nierenberg, T. Treu, X. Du, and A. Benson, Warm dark matter chills out: constraints on the halo mass function and the free-streaming length of dark matter with eight quadruple-image strong gravitational lenses, MNRAS **491**, 6077 (2020), arXiv:1908.06983 [astro-ph.CO].
[26] S. Ilić, M. Kopp, C. Skordis, and D. B. Thomas, Dark matter properties through cosmic history, Phys. Rev. D **104**, 043520 (2021), arXiv:2004.09572 [astro-ph.CO].
[27] N. Menci, An Eulerian perturbation approach to large-scale structures: extending the adhesion approximation, MNRAS **330**, 907 (2002), arXiv:astro-ph/0111228 [astro-ph].
[28] E. Kozlikin, R. Lilow, F. Fabis, and M. Bartelmann, A first comparison of Kinetic Field Theory with Eulerian Standard Perturbation Theory, J. Cosmology Astropart. Phys. **2021**, 035 (2021), arXiv:2012.05812 [astro-ph.CO].
[29] G. Knorr, Time asymptotic statistics of the Vlasov equation, J. Plasma Phys. **17**, 553 (1977).
[30] P. H. Diamond, S.-I. Itoh, and K. Itoh, *Modern Plasma Physics* (Cambridge University Press, Cambridge, 2010).
[31] M. Lesur and P. H. Diamond, Nonlinear instabilities driven by coherent phase-space structures, Phys. Rev. E **87**, 031101 (2013).
[32] M. L. Nastac, R. J. Ewart, W. Sengupta, A. A. Schekochihin, M. Barnes, and W. D. Dorland, Phase-space entropy cascade and irreversibility of stochastic heating in nearly collisionless plasma turbulence, Phys. Rev. E **109**, 065210 (2024), arXiv:2310.18211 [physics.plasm-ph].
[33] M. L. Nastac, R. J. Ewart, J. Juno, M. Barnes, and A. A. Schekochihin, Universal fluctuation spectrum of Vlasov-Poisson turbulence, forthcoming (2024).
[34] S. Konrad and M. Bartelmann, Kinetic field theory for cosmic structure formation, Rivista Nuovo Cimento **45**, 737 (2022), arXiv:2202.11077 [astro-ph.CO].
[35] Y. B. Zel'Dovich, Gravitational instability: an approximate theory for large density perturbations., A&A **500**, 13 (1970).
[36] S. F. Shandarin and Y. B. Zel'dovich, The large-scale structure of the universe: Turbulence, intermittency, structures in a self-gravitating medium, Rev. Mod. Phys. **61**, 185 (1989).
[37] J. H. Jeans, The stability of a spherical nebula, Phil. Trans. Roy. Soc. London A **199**, 1 (1902).
[38] V. I. Arnold, V. V. Kozlov, and A. I. Neishtadt, *Mathematical Aspects of Classical and Celestial Mechanics*, 3rd ed., Encyclopaedia of Mathematical Sciences, Vol. 3 (Springer-Verlag, Berlin, 2006) [Dynamical systems. III], Translated from the Russian original by E. Khukhro.
[39] J. Juno, A. Hakim, J. TenBarge, E. Shi, and W. Dorland, Discontinuous galerkin algorithms for fully kinetic plasmas, J. Comput. Phys. **353**, 110 (2018).
[40] U. Frisch, *Turbulence: The Legacy of A. N. Kolmogorov* (Cambridge University Press, Cambridge, 1995).
[41] E. Bertschinger and B. Jain, Gravitational instability of cold matter, ApJ **431**, 486 (1994), arXiv:astro-ph/9307033 [astro-ph].
[42] M. Hénon, Numerical experiments on the stability of spherical stellar systems, A&A **24**, 229 (1973).
[43] J. A. Fillmore and P. Goldreich, Self-similar gravitational collapse in an expanding universe, ApJ **281**, 1 (1984).
[44] E. Bertschinger, Self-similar secondary infall and accretion in an Einstein-de Sitter universe, ApJS **58**, 39 (1985).
[45] J. Barnes, J. Goodman, and P. Hut, Dynamical instabilities in spherical stellar systems, ApJ **300**, 112 (1986).
[46] J. Binney, Discreteness effects in cosmological N-body simulations, MNRAS **350**, 939 (2004), arXiv:astro-ph/0311155 [astro-ph].
[47] S. Colombi and J. Touma, Vlasov-Poisson in 1D: waterbags, MNRAS **441**, 2414 (2014).
[48] M. White, The Zel'dovich approximation, MNRAS **439**, 3630 (2014), arXiv:1401.5466 [astro-ph.CO].
[49] G. K. Batchelor, Small-scale variation of convected quantities like temperature in turbulent fluid. Part 1. General discussion and the case of small conductivity, J. Fluid Mech. **5**, 113 (1959).
[50] K. R. Sreenivasan, Turbulent mixing: A perspective, PNAS **116**, 18175 (2019).
[51] P. Goldreich and S. Sridhar, Toward a theory of interstellar turbulence. II. Strong Alfvenic turbulence, ApJ **438**, 763 (1995).
[52] S. V. Nazarenko and A. A. Schekochihin, Critical balance

in magnetohydrodynamic, rotating and stratified turbulence: towards a universal scaling conjecture, J. Fluid Mech. **677**, 134 (2011), arXiv:0904.3488 [physics.flu-dyn].
[53] A. A. Schekochihin, MHD turbulence: a biased review, J. Plasma Phys. **88**, 155880501 (2022), arXiv:2010.00699 [physics.plasm-ph].
[54] S. W. Hawking and G. F. R. Ellis, *The Large Scale Structure of Space-Time*, Cambridge Monographs on Mathematical Physics (Cambridge University Press, Cambridge, 1973).
[55] F. Fabis, E. Kozlikin, R. Lilow, and M. Bartelmann, Kinetic field theory: exact free evolution of Gaussian phase-space correlations, J. Stat. Mech. **4**, 043214 (2018), arXiv:1710.01611 [cond-mat.stat-mech].
[56] R. Lilow, F. Fabis, E. Kozlikin, C. Viermann, and M. Bartelmann, Resummed Kinetic Field Theory: general formalism and linear structure growth from Newtonian particle dynamics, J. Cosmology Astropart. Phys. **2019**, 001 (2019), arXiv:1809.06942 [astro-ph.CO].
[57] S. Konrad and M. Bartelmann, On the asymptotic behaviour of cosmic density-fluctuation power spectra, MNRAS **515**, 2578 (2022), arXiv:2110.07427 [astro-ph.CO].
[58] S. Konrad, Y. B. Ginat, and M. Bartelmann, On the asymptotic behaviour of cosmic density-fluctuation power spectra of cold dark matter, MNRAS **515**, 5823 (2022), arXiv:2202.08059 [astro-ph.CO].
[59] F. W. J. Olver, *Asymptotics and Special Functions*, Computer Science and Applied Mathematics (Academic Press, Cambridge, Massachussetts, 1974).
[60] N. Bleistein and R. A. Handelsman, *Asymptotic Expansions of Integrals*, 2nd ed. (Dover Publications, Inc., New York, 1986).
[61] E. W. Kolb and M. S. Turner, *The Early Universe* (CRC Press, Taylor & Francis, London, 1990).
[62] R. J. Ewart, M. L. Nastac, P. J. Bilbao, T. Silva, L. O. Silva, and A. A. Schekochihin, Relaxation to universal non-Maxwellian equilibria in a collisionless plasma, (2024), arXiv:2409.01742 [physics.plasm-ph].
[63] W. H. Press and P. Schechter, Formation of galaxies and clusters of galaxies by self-similar gravitational condensation, ApJ **187**, 425 (1974).
[64] C.-P. Ma and J. N. Fry, Deriving the nonlinear cosmological power spectrum and bispectrum from analytic dark matter halo profiles and mass functions, ApJ **543**, 503 (2000), arXiv:astro-ph/0003343 [astro-ph].
[65] M. Asgari, A. J. Mead, and C. Heymans, The halo model for cosmology: a pedagogical review, The Open Journal of Astrophysics **6**, 10.21105/astro.2303.08752 (2023).
[66] A. Kolmogorov, The local structure of turbulence in incompressible viscous fluid for very large Reynolds' numbers, Akad. Nauk SSSR Doklady **30**, 301 (1941).
[67] J. P. Dougherty, Model Fokker-Planck equation for a plasma and its solution, Phys. Fluids **7**, 1788 (1964).
[68] A. Hakim, M. Francisquez, J. Juno, and G. W. Hammett, Conservative discontinuous Galerkin schemes for nonlinear Dougherty-Fokker-Planck collision operators, J. Plasma Phys. **86**, 905860403 (2020), arXiv:1903.08062 [physics.comp-ph].
[69] T. Tatsuno, W. Dorland, A. A. Schekochihin, G. G. Plunk, M. Barnes, S. C. Cowley, and G. G. Howes, Nonlinear phase mixing and phase-space cascade of entropy in gyrokinetic plasma turbulence, Phys. Rev. Lett. **103**, 015003 (2009), arXiv:0811.2538 [physics.plasm-ph].
[70] M. Davis and P. J. E. Peebles, On the integration of the BBGKY equations for the development of strongly nonlinear clustering in an expanding universe, ApJS **34**, 425 (1977).

Appendix A: Simulation methods

Here we describe the details of the Vlasov–Poisson simulations whose results were shown in §II and figures 1–5. The simulations were conducted using the Gkeyll code[12], which is an Eulerian solver in phase-space, originally designed for the Vlasov-Maxwell system. Gkeyll uses a discontinuous Galerkin algorithm to compute distribution functions [39]. We set the vacuum permittivity $\varepsilon_0 = -1$, which turns electrostatic interactions into gravitational ones—and, for unit particle charges, it is equivalent to choosing units such that $4\pi G = 1$. Our simulations were 1D in position space (so the phase-space is 2D), with periodic spatial boundary conditions; the units of length were such that the box size was $L = 2\pi$. We always took the particle mass to be unity, which implied that the system's total mass was normalised to $M = 2\pi$. These three choices specify the units for the simulations, and are equivalent to choosing $\tau_0 = k_0 = v_0 = 1$ (these are defined in §II A). Thus, the outer scale $k_{\rm nl} \sim k_0$ is of order unity.

[12] https://gkeyll.readthedocs.io/

1. Initial conditions

We simulated a single-species distribution function, with the initial condition (4), where we chose $v_{\rm th} = 0.005v_0$ or $0.01v_0$, and

$$\rho_{\rm in}(x)\frac{L}{M} = 1 - \sum_{n=1}^{5} \frac{a_n}{k_n} \sin\left(k_n x + \phi_n\right), \tag{A1}$$

$$u_{\rm in}(x) = v_0 \sum_{n=1}^{5} b_n \cos\left(k_n x + \phi_n\right), \tag{A2}$$

where $\phi_n \in [0, 2\pi]$ are random phases, a_n/k_0 were uniformly sampled from the interval $[0, 0.2]$, and $k_n = k_0 U_n$, with $U_n \in \{1, 2, \dots, 10\}$—a uniformly distributed random integer. For runs with one initial stream (figure 5), $b_n = a_n/k_0$, while for runs with three initial streams (figure 1), we set $b_n = 0.05a_n/k_0$ and tripled the initial conditions (4) by shifting $u_{\rm in}(x)$ by $\pm 2v_0$ and duplicating for two additional streams. We ensured that the resolution was sufficiently fine for a Maxwellian of width $v_{\rm th}$ still to be resolved: we used an $N_x \times N_v$ phase-space grid with $N_x = N_v = 4032$ for the single-stream initial condition and $N_x = N_v = 7680$ for the multi-stream one.

Because of numerical errors, f can become slightly negative in very small and isolated areas. This is because the discontinuous Galerkin algorithm that Gkeyll uses is not positivity-preserving, and errors arise from over-shooting due to large derivatives. These problems are somewhat alleviated by the small collision operator (see appendix A 2), which smooths large velocity derivatives. In any case, these regions do not cause the simulation to become unstable because they are isolated and $|f|$ is still very small there; the total mass occupied by negative f is $\iint f\Theta(-f){\rm d}x{\rm d}v < 10^{-5}M$ for the simulation in figure 1 and $< 0.02M$ for figure 5. Therefore, this does not invalidate any of our conclusions in this paper, and we have set $f \mapsto f\Theta(f)$ for the purpose of plotting figures 1 and 5.

2. Collisions

One would like to run collisionless simulations, but the finite resolution of any numerical representation of phase-space induces an unavoidable effective collisionality due to the grid. Real dark matter also has a finite number of particles—and finite-N effects also induce a effective collisionality. To control the cut-off scale in velocity space and the dissipation of C_2, we added a weak collision operator $\left(\partial f/\partial t\right)_c \propto \nu$, where ν is the collision frequency. The collision operator used was the Dougherty collision operator—a type of Fokker-Planck operator in velocity space [67, 68]:

$$\left(\frac{\partial f}{\partial t}\right)_c = \nu \frac{\partial}{\partial v}\left[\left(v - u\right) f + v_t^2 \frac{\partial f}{\partial v}\right], \tag{A3}$$

where $u(x)$ is first velocity moment of f and $v_t^2 \equiv \int {\rm d}v\ v^2 f/\rho$. Details (and the definition of v_t) can be found in [68], which describes the implementation of equation (A3) in Gkeyll. The inevitability of some collisions—whether due to the grid, finite-N effects, or a collision operator—implies the existence of another scale in the problem. We used the Dorland number [69] ${\rm Do} \equiv (\nu\tau_0)^{-1} = 10^6$ for all runs. The collision time scale is [32, 68]

$$\tau_\nu \sim \frac{1}{\nu s^2 v_{\rm rms}^2}, \tag{A4}$$

where $v_{\rm rms}^2$ is the second velocity cumulant. The velocity scale where collisions become competitive with the cascade rate ($\tau_\nu \sim \tau_{\rm g}$) is, therefore,

$$u_\nu = \sqrt{\nu\tau_{\rm g}}\, v_{\rm rms}. \tag{A5}$$

By critical balance (see §II C), this also gives a length scale

$$l_\nu = \sqrt{\nu\tau_{\rm g}}\, L \sim \sqrt{\nu\tau_{\rm g}}\, k_{\rm nl}^{-1}, \tag{A6}$$

which is the collisional cut-off in position space. For a sufficiently small ν, one can have[13]

$$k_{\rm nl} \ll k \ll k_{\rm c}(v_{\rm th}^{-1}) \ll l_\nu^{-1}, \tag{A7}$$

$$v_{\rm rms}^{-1} \ll s \ll v_{\rm th}^{-1} \ll u_\nu^{-1}, \tag{A8}$$

[13] In reality, it might be that $u_\nu \gtrsim v_{\rm th}$, for both parameters depend sensitively on the nature of the constituent particle(s) of dark matter. But as long as $1/s$ is larger than both, none of the conclusions of the this paper are invalidated by this possibility.

where $k_{\mathrm{c}}(s)$ is defined by equation (29). While this hierarchy justifies neglecting collisions in this paper, as phase-space structure cascades to ever smaller scales by (18), the collisional scales, where it is erased, are eventually reached [32, 33]. Likewise, in reality, a particle-noise floor $\hat{F} \approx M^2/N$ (which is not present in our simulations) is eventually reached, too. Determining whether this occurs on scales smaller or larger than l_ν is deferred to future work [cf. 33].

3. Phase-space power spectra

In the 1D set-up described above, our prediction (55) for the phase-space power spectrum becomes

$$\hat{F}(k,s) \sim \begin{cases} F_1 k^{-1}, & \text{if } \max\left\{k_{\mathrm{nl}}, k_{\mathrm{c}}(s)\right\} \ll k \ll \min\left\{k_{\mathrm{c}}(v_{\mathrm{th}}^{-1}), l_\nu^{-1}\right\}, \\ F_2 s^{-1}, & \text{if } \max\left\{v_{\mathrm{rms}}^{-1}, s_{\mathrm{c}}(k)\right\} \ll s \ll \min\left\{v_{\mathrm{th}}^{-1}, u_\nu^{-1}\right\}. \end{cases} \tag{A9}$$

We are not concerned here with the asymptotics of the power spectrum $\hat{F}$, defined in equation (10), at values of s and k above v_{th}^{-1} and $k_{\mathrm{c}}(v_{\mathrm{th}}^{-1})$, respectively, but below the collisional cut-offs u_ν^{-1} and l_ν^{-1} (see equations (A5) and (A6)). In this region, gravitational collapse halts by equation (36), but if one extrapolates the findings of [33]—who explored turbulence in electrostatic plasmas where v_{th} is of order unity—from electrostatics to gravity (ignoring any possible differences), then one should have, assuming $v_{\mathrm{th}} \gg u_\nu$,

$$\hat{F}(k,s) \sim \begin{cases} F_1 k^{-1}, & \text{if } \max\left\{k_{\mathrm{nl}}, k_{\mathrm{c}}(s)\right\} \ll k \ll k_{\mathrm{c}}(v_{\mathrm{th}}^{-1}), \\ F_2 s^{-1}, & \text{if } \max\left\{v_{\mathrm{rms}}^{-1}, s_{\mathrm{c}}(k)\right\} \ll s \ll v_{\mathrm{th}}^{-1}, \\ F_3 k^{-2}, & \text{if } \max\left\{k_{\mathrm{nl}}, k_{\mathrm{c}}(s), k_{\mathrm{c}}(v_{\mathrm{th}}^{-1})\right\} \ll k \ll l_\nu^{-1}, \\ F_4 s^{-2}, & \text{if } \max\left\{v_{\mathrm{rms}}^{-1}, s_{\mathrm{c}}(k), v_{\mathrm{th}}^{-1}\right\} \ll s \ll u_\nu^{-1}. \end{cases} \tag{A10}$$

The spectrum is truncated exponentially at $k > l_\nu^{-1}$ or $s > u_\nu^{-1}$ (or submerged into particle noise as discussed in [33]).

Figure 1 shows an example of the time evolution of a system that started with three streams with $v_{\mathrm{th}} = 0.005 v_0$ and had $\mathrm{Do} = 10^6$. This corresponds to $u_\nu \sim 10^{-3} v_0$, so scales are separated in the way that we have assumed. The grid size is $N_x = N_v = 7680$ and the velocity box size is $12 v_0$, so the Nyquist velocity scale is $s_{\mathrm{Ny}} \approx 2011 v_0^{-1}$. Figure 1 shows that the streams collapse quickly, by rotating and twisting in phase-space. In figure 3, we show the evolution of the power spectra for this simulation.

Appendix B: Estimates for the source

In this appendix, we show that the scaling (41) of the source term, argued for physically in §II E 2, is *a posteriori* consistent with the phase-space power spectrum in equation (55) (derived from it) and the formal definition (19) of the source term.

Consider the integral (35). As argued in §II E 2, the magnitude of the contribution of $\int \hat{S}(\mathbf{k},\mathbf{s})\mathrm{d}^d s$ at each scale k is proportional to the inverse collapse time τ_{g}^{-1}, which gave $\mathcal{F}^{\mathbf{k}} \sim \varepsilon k^d/\tau_{\mathrm{g}}$. Let us verify this by examining the scalings of each term in equation (19) with k and s. By critical balance (§II C) and equations (19) and (35), the integrated source is

$$\mathcal{F}^{\mathbf{k}} = (2\pi)^{-2d} \int^k \mathrm{d}^d k' \int \mathrm{d}^d s \; \hat{S}(\mathbf{k}',\mathbf{s}) \sim \frac{V}{\tau_{\mathrm{nl}}} \int^k \mathrm{d}^d k' \int \mathrm{d}^d s \; \left\langle \left| \hat{\overline{f}}(\mathbf{s}) \hat{f}^*(\mathbf{k}',\mathbf{s}) \right| \right\rangle, \tag{B1}$$

where we have already carried out the Fourier transform in equation (19). Crucially, the (real-space) product $\mathbf{g}\cdot\partial/\partial\mathbf{v} \sim \tau_{\mathrm{nl}}^{-1}$ in (19) is proportional to the (inverse) collapse time, τ_{g}^{-1}, a constant (cf. §§II C and II E 2). Furthermore, its sign must also be independent of scale: as long as $\max\left\{s, s_{\mathrm{c}}(k)\right\} v_{\mathrm{th}} \ll 1$, nothing in equation (36) depends on scale, since in this limit, the collapse is independent of scale.

To estimate (B1), we use the following estimates:

$$\left| \hat{f}^*(\mathbf{k}',\mathbf{s}) \right|^2 \sim \hat{F}(k',s), \tag{B2}$$

$$\left| \hat{\overline{f}}(\mathbf{s}) \right|^2 \sim \frac{1}{V^2} \hat{F}(0,s); \tag{B3}$$

these follow from equations (6) and (10). At each $\mathbf{k}'$, the $\mathbf{s}$ integral in (B1) has one contribution from $0 < s < s_{\mathrm{c}}(k')$, where we can approximate $\hat{F}(k',s) \sim \hat{F}(k',0)$, for $k' \ll k$, by critical balance; for the other contribution, from

$s \gg s_c(k')$, we approximate $\hat{F}(k', s) \sim \hat{F}(0, s)$. Hence

$$\mathcal{F}^{\mathbf{k}} \sim \frac{1}{\tau_g} \int^k d^d k' \int^{s_c(k')} d^d s \sqrt{\hat{F}(k', 0)\hat{F}(0, s)} + \frac{1}{\tau_g} \int^k d^d k' \int_{s_c(k')} d^d s \; \hat{F}(0, s). \tag{B4}$$

Let us parameterise, as in equation (43), $\hat{F}(k, 0) \sim F_1 k^\gamma$ and $\hat{F}(0, s) \sim F_2 s^\delta$, where both γ and δ are negative; critical balance (equations (26) and (29)) gives $\gamma = \delta$, as indeed found in §II F. Then, the integrations result in

$$\mathcal{F}^{\mathbf{k}} \sim \frac{\sqrt{F_1 F_2}}{\tau_g \kappa^{(d+\delta)/2}} k^{2d+(\gamma+\delta)/2} + \frac{F_2}{\tau_g \kappa^{d+\delta}} k^{2d+\delta}, \tag{B5}$$

which scales like $k^{2d+\delta/2+\max\{\gamma,\delta\}/2}$. If $\max\{\gamma, \delta\} = -d$, then $\mathcal{F}^{\mathbf{k}} \propto k^d$. A similar calculation yields the analogous result $\mathcal{F}^{\mathbf{s}} \propto s^d$ at large s ($\mathcal{F}^{\mathbf{s}}$ is defined in equation (34)).

The scaling $\mathcal{F}^{\mathbf{k}} \propto k^d$ is the only one consistent with sourcing by gravitational collapse (§II E 2). But we know from equation (55) and from the results of §III that the density power spectrum of cold dark matter assumes its k^{-d} shape after roughly one critical-balance time, so indeed $\max\{\gamma, \delta\} = -d$, and $\mathcal{F}^{\mathbf{k}} \propto k^d$ is consistent with equations (19) and (55), as required. This power spectrum in turn keeps producing a source with just the right scaling, to ensure that $\hat{F}$ retains the same k^{-d} power law. This is an *a posteriori* justification for $\iint \hat{S} \sim \varepsilon k^d/\tau_g$.

Appendix C: Einstein-de-Sitter universe

Although the analysis of §II, resulting in the prediction (55) for the phase-space power spectrum, is valid for ΛCDM, and in general for any cosmology where there is a separation of scales between $\mathcal{H}/c$ and the non-linear scale $k_{\rm nl}$, in the special case of an Einstein-de-Sitter (EdS) cosmology, there exists a scale invariance (see, e.g., [1, 70]) that provides additional insight, which we sketch in this appendix. This allows us to propose a global analogue of the local analysis of the main text. We focus on the case of $d = 3$ for simplicity.

In an EdS space-time, the Vlasov–Poisson system is invariant under the transformation

$$(t, \mathbf{x}, \mathbf{v}, f) \mapsto (\lambda t, \lambda^\zeta \mathbf{x}, \lambda^{\zeta-1/3}\mathbf{v}, \lambda^{3\zeta+1} f), \tag{C1}$$

where ζ is fixed by assuming the scaling of the density power spectrum. This means that for every (inverse) length scale k, there exists a (cosmic) time scale $t \propto k^{1/\zeta}$. Thus, one obtains a (conformal) time scale, which is just the collapse conformal time in the standard spherical-collapse model,

$$\tau_g(k) \equiv \frac{2}{H_0} \left(\frac{\pi}{2H_0 t_0}\right)^{1/3} \left(\frac{3}{5\delta_{\rm rms}(k)}\right)^{1/2}, \tag{C2}$$

where H_0 and t_0 are the present-day Hubble constant and cosmic time, and $\delta_{\rm rms}(k)$ is the root-mean-square amplitude of the dark-matter density fluctuation at scale k, normalised by the EdS growth factor $D_+(t_{\rm initial}) = a(t_{\rm initial})$. The time scale in equation (C2) turns out to be nothing but the standard gravitational time scale $\tau_g \sim 1/\sqrt{G\rho}$, where $\rho \sim \delta_{\rm rms}\rho_{\rm crit}$, with the critical density defined by $\rho_{\rm crit} = 3H_0^2/(8\pi G)$. Relating $\delta_{\rm rms}(k)$ to the over-density power spectrum $P_\delta(k) = P_0(k/k_{\rm p})^n$ for some constant $k_{\rm p}$, via $\delta_{\rm rms}(k) = \sqrt{k^3 P_\delta(k)}$, yields

$$\tau_g \sim \frac{2\sqrt{3}}{\sqrt{5}H_0} \left(\frac{\pi}{2H_0 t_0}\right)^{1/3} \left(\frac{k_{\rm p}^n}{P_0 k^{3+n}}\right)^{1/4}. \tag{C3}$$

By critical balance (§II C), the collapse time scale τ_g and the linear, phase-mixing time τ_l, must be the same, whence

$$s_c(k) \propto k^{(1-n)/4}. \tag{C4}$$

Consequently, the analogue of equation (48) here would imply

$$k^{4+\gamma} \sim s_c(k). \tag{C5}$$

Using equation (C3), one concludes from the analogue of equation (48), that the exponent γ of the density power spectrum in (43) is $\gamma = -(15 + n)/4$. In a steady state, the power spectrum must be invariant under the collapse process, and so in a self-sustaining scenario, $\gamma = n$, which is solved by $n = -3$. This is unsurprising, because an EdS space-time is just a matter-dominated FLRW space-time; as this paper is focused on small scales, locally, everything is matter dominated.

Appendix D: Initial momentum correlations

The purpose of this appendix is to prove, for §III, that $\Sigma \sim Aq^2$ as $q \to 0$, and that $\Sigma\mathbf{a} = O(q)$, which are defined in equation (58). What follows holds for $v_{\rm th} = 0$.

Consider equation (58). Since $2\mathbf{p}_1 = \mathbf{p} + 2\mathbf{P}$ and $2\mathbf{p}_2 = 2\mathbf{P} - \mathbf{p}$, the part involving $\mathbf{a}$ comes solely from the cross-correlations between $\mathbf{p}_1$, $\mathbf{p}_2$ and $(\mathbf{p}_3, \dots, \mathbf{p}_N)$, *viz.*,

$$\begin{aligned}\mathbf{p}_1^T (C_{pp}^N)^{-1}_{1n}\mathbf{p}_n + \mathbf{p}_2^T (C_{pp}^N)^{-1}_{2n}\mathbf{p}_n &= \mathbf{P}\left[(C_{pp}^N)^{-1}_{1n} + (C_{pp}^N)^{-1}_{2n}\right]\mathbf{p}_n + \frac{1}{2}\mathbf{p}\left[(C_{pp}^N)^{-1}_{1n} - (C_{pp}^N)^{-1}_{2n}\right]\mathbf{p}_n \\ &+ \mathbf{P}\left[(C_{pp}^N)^{-1}_{11} - (C_{pp}^N)^{-1}_{22}\right]\mathbf{p} + 2\mathbf{P}\left[(C_{pp}^N)^{-1}_{11} + (C_{pp}^N)^{-1}_{21} + (C_{pp}^N)^{-1}_{12} + (C_{pp}^N)^{-1}_{22}\right]\mathbf{P},\end{aligned} \tag{D1}$$

where $(C_{pp}^N)^{-1}_{mn}$ specifies the sub-matrix of $(C_{pp}^N)^{-1}$ that pertains to particles m and n, and the Einstein summation convention is used.

Now, $(C_{pp}^N)^{-1}_{mn}$ is a function of the positions of the particles only. When $q \equiv |\mathbf{q}_1 - \mathbf{q}_2| \to 0$, $\mathbf{q}_1$ and $\mathbf{q}_2$ both tend to $\mathbf{Q} = (\mathbf{q}_1 + \mathbf{q}_2)/2$, and, by the indistinguishability of the constituent particles,

$$(C_{pp}^N)^{-1}_{1n}(\mathbf{q}_1 = \mathbf{Q}, \mathbf{q}_2 = \mathbf{Q}, \mathbf{q}_3, \dots, \mathbf{q}_N) = (C_{pp}^N)^{-1}_{2n}(\mathbf{q}_1 = \mathbf{Q}, \mathbf{q}_2 = \mathbf{Q}, \mathbf{q}_3, \dots, \mathbf{q}_N), \tag{D2}$$

$$(C_{pp}^N)^{-1}_{11}(\mathbf{q}_1 = \mathbf{Q}, \mathbf{q}_2 = \mathbf{Q}, \mathbf{q}_3, \dots, \mathbf{q}_N) = (C_{pp}^N)^{-1}_{22}(\mathbf{q}_1 = \mathbf{Q}, \mathbf{q}_2 = \mathbf{Q}, \mathbf{q}_3, \dots, \mathbf{q}_N). \tag{D3}$$

Consequently, for $q \to 0$ and finite $\mathbf{Q}, \mathbf{q}_3, \dots, \mathbf{q}_N, \mathbf{P}, \mathbf{p}_3, \dots, \mathbf{p}_N$, we have

$$\left[(C_{pp}^N)^{-1}_{1n} - (C_{pp}^N)^{-1}_{2n}\right] = O(q), \tag{D4}$$

$$\left[(C_{pp}^N)^{-1}_{11} - (C_{pp}^N)^{-1}_{22}\right] = O(q). \tag{D5}$$

Therefore, in the limit $q \to 0$, the matrix $(C_{pp}^N)^{-1}$, *qua* a linear operator in momentum space, splits into a block matrix in the basis $(\mathbf{p}, \mathbf{P}, \mathbf{p}_3, \dots, \mathbf{p}_N)$, with one block for $\mathbf{p}$ and another for $(\mathbf{P}, \mathbf{p}_3, \dots, \mathbf{p}_N)$. Hence its inverse C_{pp}^N is also a block matrix in this basis. Namely,

$$C_{pp}^N = \left(\begin{array}{c|cc} (C_{pp}^N)_{11} + (C_{pp}^N)_{22} - (C_{pp}^N)_{12} - (C_{pp}^N)_{21} & 0 & (C_{pp}^N)_{1n} - (C_{pp}^N)_{2n} \\ \hline 0 & (C_{pp}^N)_{11} + (C_{pp}^N)_{22} + (C_{pp}^N)_{12} + (C_{pp}^N)_{21} & (C_{pp}^N)_{1n} + (C_{pp}^N)_{2n} \\ (C_{pp}^N)_{1n} - (C_{pp}^N)_{2n} & (C_{pp}^N)_{1n} + (C_{pp}^N)_{2n} & (C_{pp}^N)_{nm} \end{array}\right). \tag{D6}$$

The matrix Σ^{-1} is the top-left block of $(C_{pp}^N)^{-1}$:

$$\Sigma = \left[(C_{pp}^N)_{11} + (C_{pp}^N)_{22} - (C_{pp}^N)_{12} - (C_{pp}^N)_{21}\right] - \left[(C_{pp}^N)_{1n} - (C_{pp}^N)_{2n}\right](C_{pp}^N)^{-1}_{nm}\left[(C_{pp}^N)_{1m} - (C_{pp}^N)_{2m}\right]. \tag{D7}$$

This, in conjunction with results on the initial correlation sub-matrix of C_{pp}^N pertaining to particles 1 and 2 [34, 57, 58] to the effect that when $v_{\rm th} = 0$, $(C_{pp}^N)_{11} + (C_{pp}^N)_{22} - (C_{pp}^N)_{12} - (C_{pp}^N)_{21} = O(q^2)$ (with the coefficient determined by the cosmology, which we take to be standard ΛCDM), implies that indeed $\Sigma \sim Aq^2$ as $q \to 0$. Besides, the components a_k of $\mathbf{a}$ satisfy

$$\mathrm{a}_k \propto -\left(\Sigma^{-1}\right)_{kn}\left[(C_{pp}^N)_{1n} - (C_{pp}^N)_{2n}\right](C_{pp}^N)^{-1}_{nm}\mathbf{p}_m, \tag{D8}$$

so

$$\Sigma\mathbf{a} \sim \left[(C_{pp}^N)_{1n} - (C_{pp}^N)_{2n}\right] = O(q). \tag{D9}$$

Additionally, the results of [34, §3.1.2], in conjunction with equation (D6), imply that the entire C_{pp}^N matrix tends to a constant as $q \to \infty$, which means that Σ, $\mathbf{a}$ and B become $O(1)$ in that limit, too.