

The role of energy shear in the collapse of protohaloes

Marcello Musso^{1,*}  and Ravi K. Sheth^{2,3,*} 

¹ Departamento de Física Fundamental and IUFFyM, Universidad de Salamanca, Plaza de la Merced s/n, 37008 Salamanca, Spain

² Center for Particle Cosmology, University of Pennsylvania, 209 S. 33rd St., Philadelphia, PA 19104, USA

³ The Abdus Salam International Centre for Theoretical Physics, Strada Costiera 11, 34151 Trieste, Italy

Received 24 March 2025 / Accepted 30 November 2025

ABSTRACT

Dark matter haloes form from the collapse of matter around special positions in the initial field, those where the local matter flows converge to a point. For such a triaxial collapse to take place, the energy shear tensor (the source of the evolution of the inertia tensor) must be positive definite. It has been shown that this is indeed the case for the energy shear tensor of the vast majority of protohaloes. At generic positions in a Gaussian random field, the trace and traceless parts of the tensor are independent of one another. Here we show that, on the contrary, in positive definite matrices they correlate strongly, and these correlations are very similar to those exhibited by protohaloes. Moreover, while positive-definiteness ensures that an object will collapse, it does not specify when. Previous work has shown that the trace of the energy tensor (the energy overdensity) exhibits significant scatter in its values, but must lie above a critical ‘threshold’ value for the halo to collapse by today. We show that suitable combinations of the eigenvalues of the traceless part are able to explain a substantial part of the scatter of the trace. These variables provide an efficient way to parametrise the initial value of the energy overdensity, allowing us to formulate an educated guess for the threshold of collapse. We validate our ansatz by measuring the distribution of several secondary properties of protohaloes, finding good agreement with our analytical predictions.

Key words. large-scale structure of Universe

1. Introduction

The assembly, abundance and clustering of dark matter haloes play an important role both in galaxy formation and in the ongoing efforts to constrain cosmological parameters, where it is crucial to model accurately how haloes and galaxies trace the matter density field. While the latter is continuous, though, at least above some scale, the halo centres of mass define a point process at a given time, and the same is true at the initial time for the centres of mass of the protohalo patches from which they formed [Bardeen et al. \(1986\)](#). Therefore, models of halo collapse and formation seek to identify what makes these locations special in the initial fluctuation field. Such studies make two types of choices: what the ‘field’ is, and what the ‘properties’ of this field are.

Following [Gunn & Gott \(1972\)](#), who described halo formation as the collapse of a homogeneous sphere, most of the literature on the subject has assumed that the initial matter overdensity field is the one of interest. However, haloes are not spherical. Thus, some early works suggested that the ‘shape tensor’ – identified with the Hessian of the matter density field, that is, the 3×3 matrix of the second spatial gradients – provides useful additional information about departures from sphericity ([Bardeen et al. 1986](#); [van de Weygaert & Bertschinger 1996](#)).

Other authors focused on a different 3×3 matrix: the ‘deformation tensor’ – constructed from the volume average of the second gradients of the gravitational potential – noting that it is dynamically more interesting ([Doroshkevich 1970](#); [Bond & Myers 1996](#)). Indeed, in the homogeneous ellipsoidal collapse model studied by [Bond & Myers \(1996\)](#), an initially spherical region is deformed by the surrounding shear field:

hence the name ‘deformation’ tensor (routinely used in fluid dynamics, although usually referring to infinitesimal volumes, not to the average over macroscopic regions). The mean matter overdensity, thanks to Poisson’s equation, is the trace of the deformation tensor; and since the trace is the sum of the three eigenvalues, this raises the question whether the 3 eigenvalues (or their linear combinations) themselves play an important role, or if only their sum matters.

Any 3×3 matrix also has three ‘rotational invariants’. These are the coefficients of its characteristic polynomial, whose roots are the eigenvalues, and provide an alternative description of the same information. Since the trace is just one of the three, it is natural to ask what role the other two, which are non-linear combinations of the eigenvalues, play. This is particularly interesting in view of the fact that rotational invariants have been used to quantify how protohaloes correlate with their environment ([Desjacques 2013](#); [Lazeyras et al. 2016](#); [Lazeyras & Schmidt 2018](#)).

Indeed, the second order invariant of the deformation tensor – the amplitude of the traceless shear – matters for halo formation: protohaloes with larger shear tend to be more overdense ([Sheth et al. 2001](#); [Despali et al. 2013](#)), and this sources a correlation with the environment on larger scales ([Sheth et al. 2013](#); [Modi et al. 2017](#); [Lazeyras & Schmidt 2018](#)). This is slightly surprising because, in a Gaussian random field, the two invariants are statistically independent ([Sheth & Tormen 2002](#)). Therefore, we want to investigate what aspect of halo formation induces the measured correlation between overdensity and shear, since this another manifestation of what makes protohalo positions special.

Recently, it has been highlighted that, as a ‘field’, the initial energy overdensity is more interesting than the initial matter overdensity ([Musso & Sheth 2021](#)). Indeed, the energy shear

* Corresponding authors: mmusso@usal.es, shethrk@upenn.edu

tensor is more directly related to the physics of collapse than is the deformation tensor (Musso et al. 2024). This implies that, unlike the deformation tensor, the energy shear is almost always positive definite. Moreover, just as for the deformation tensor and even more strongly, the first and second invariants of the energy shear are positively correlated.

With this in mind, our study has two main goals. First, does the third invariant (of the energy shear tensor) also matter? And second, can the correlations between the invariants of the protohalo patches be understood as simply arising from the positivity constraint?

In the next section, we define the three invariants, before showing that all three play an important role in characterising the properties of protohalo patches. Then we show that requiring positivity induces correlations between the invariants, which are qualitatively similar to those exhibited by protohaloes. While positivity implies that an object will collapse, it does not determine when the collapse is complete: Section 3 discusses how collapse by a specified time can be associated with a threshold in the energy overdensity. It then explores if this threshold depends on other invariants (than the trace) as well. A final section summarises our conclusions.

2. Rotational invariants of protohaloes

The energy overdensity tensor of a generic comoving volume V , or energy shear, is the 3×3 matrix

$$u_{ij} \equiv \frac{3}{I} \int_V d\mathbf{r} \rho(\mathbf{r}, t) (r_i - r_{\text{cm},i}) (\nabla_j \phi - [\nabla_j \phi]_{\text{cm}}) \quad (1)$$

as defined by Musso et al. (2024), where ρ is the matter density, ϕ is the potential energy perturbation (related to the matter density contrast $\delta = \rho/\bar{\rho} - 1$ by the Poisson equation $\nabla^2 \phi = \delta$), $\mathbf{r}_{\text{cm}} \equiv (1/M) \int_V d\mathbf{r} \rho(\mathbf{r}, t) \mathbf{r}$ and $[\nabla \phi]_{\text{cm}} \equiv (1/M) \int_V d\mathbf{r} \rho(\mathbf{r}, t) \nabla \phi$ are the centre of mass position and acceleration, and $I \equiv \int_V d\mathbf{r} \rho(\mathbf{r}, t) |\mathbf{r} - \mathbf{r}_{\text{cm}}|^2$ is the moment of inertia.

At very early times, when ρ is nearly uniform and ϕ is a nearly Gaussian field, measuring u_{ij} at generic positions returns a Gaussian random matrix. If V is a randomly placed sphere of radius R , and only in this case, the covariance of the energy shear can be expressed as

$$\langle u_{ij} u_{kl} \rangle = \frac{\delta_{ik} \delta_{jl} + \delta_{il} \delta_{jk} + \delta_{ij} \delta_{kl}}{15} \sigma_{02}^2, \quad (2)$$

where the Gaussian moment

$$\sigma_{02}^2(R) \equiv \int \frac{d\mathbf{k}}{(2\pi)^3} \delta(\mathbf{k}) \frac{15 j_2(kR)}{(kR)^2} \quad (3)$$

is the variance of the trace of u_{ij} , or energy overdensity (Musso & Sheth 2021). All higher cumulants of the distribution vanish. Since $M = (4\pi/3) \bar{\rho} R^3$, σ_{02} encodes the mass dependence of the distribution.

However, as pointed out recently by Musso et al. (2024), u_{ij} also sources the evolution of the inertia tensor I_{ij} of the body. Therefore, if V is a protohalo, the energy shear (more precisely, its symmetric part $u_{ij} + u_{ji}$) must have three positive eigenvalues for the volume to collapse along three axes. In other words, measuring u_{ij} in protohaloes no longer yields a Gaussian random matrix, but a positive definite one.

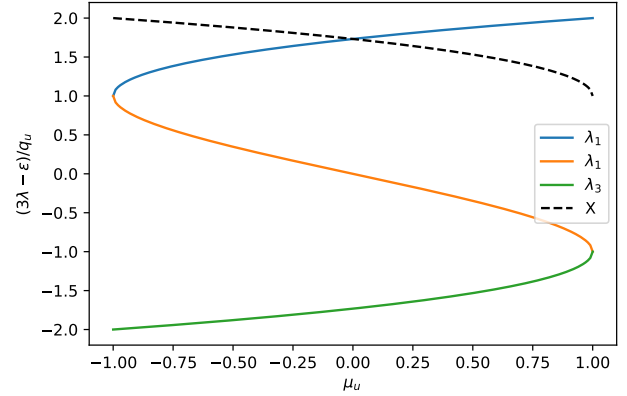


Fig. 1. Ratio of traceless eigenvalues to $q/3$ as a function of μ (equation 7). Also shown (dashed line) is the function $X(\mu)$ defined in equation (10), marking the lower limit of ϵ/q .

2.1. Rotational invariants of positive definite matrices

Regardless of its signature, the symmetric part $u_{ij} + u_{ji}$ of the energy shear has three real eigenvalues $\lambda_1 > \lambda_2 > \lambda_3$, a priori not necessarily positive, in terms of which one can construct the three rotational invariants

$$\begin{aligned} \epsilon &\equiv \sum_i \lambda_i, & q^2 &\equiv \frac{3}{2} \sum_i (\lambda_i - \epsilon/3)^2, & \text{and} \\ U^3 &\equiv \sum_i (\lambda_i - \epsilon/3)^3 = 3 \prod_i (\lambda_i - \epsilon/3). \end{aligned} \quad (4)$$

The trace $\epsilon = \text{tr}(\mathbf{u})$ is the energy overdensity; in terms of $\bar{u}_{ij} \equiv u_{ij} - \epsilon \delta_{ij}/3$, the traceless part of the matrix, $q^2 = 3 \text{tr}(\bar{\mathbf{u}}^2)/2$ is the traceless shear amplitude; and $U^3 = \text{tr}(\bar{\mathbf{u}}^3) = 3 \det(\bar{\mathbf{u}}^3)$ is (thrice) the traceless determinant.

The equation for the eigenvalues can be conveniently written in the form of a depressed cubic as

$$[(3\lambda - \epsilon)/q]^3 - 3[(3\lambda - \epsilon)/q] - 2\mu = 0, \quad (5)$$

where

$$\mu \equiv (9/2) (U^3/q^3); \quad (6)$$

for this equation to admit three real solutions, as it should be since the matrix is symmetric, the constraint $\mu^2 < 1$ must be satisfied, or equivalently $-1 \leq \mu \leq 1$. Clearly, one also has $q \equiv \sqrt{q^2} \geq 0$. The roots of the cubic equation are then

$$\lambda_k = \frac{\epsilon}{3} + \frac{2q}{3} \cos \left[\frac{1}{3} \arccos(\mu) - \frac{2\pi}{3}(k-1) \right], \quad (7)$$

which expresses the three eigenvalues in terms of the three rotational invariants. Configurations with $\mu \simeq 1$ have $\lambda_1 > \lambda_2 \simeq \lambda_3$, whereas $\mu \simeq -1$ implies $\lambda_1 \simeq \lambda_2 > \lambda_3$, and $\mu \simeq 0$ implies $\lambda_2 \simeq \epsilon/3$. Figure 1 shows the three eigenvalues as a function of μ . Evidently, $3(\lambda_1 - \lambda_3)/q \approx 3$ for all values of μ , meaning that $\lambda_1 - \lambda_3 \approx q$. Hence, the ellipticity $e_u \equiv (\lambda_1 - \lambda_3)/2\epsilon \approx q/2\epsilon$.

Regardless of the sign of the eigenvalues, the rotational invariants always satisfy the relations

$$\epsilon^2 - q^2 = 3(\lambda_1 \lambda_2 + \lambda_1 \lambda_3 + \lambda_2 \lambda_3) \quad \text{and} \quad (8)$$

$$2\mu + \frac{\epsilon}{q} \left(\frac{\epsilon^2}{q^2} - 3 \right) = 27 \frac{\det(\mathbf{u})}{q^3} = 27 \frac{\lambda_1 \lambda_2 \lambda_3}{q^3}. \quad (9)$$

If however u_{ij} is positive definite (that is, if all its eigenvalues are positive), the right-hand sides of equations (8) and (9) are both positive; so $\epsilon > q$ and $(\epsilon/q)^3 - 3(\epsilon/q) + 2\mu > 0$. The equation associated with the latter inequality is the same as the equation for $(\epsilon - 3\lambda)/q$, and its roots can be obtained from equation (7).

The only root larger than 1 (required since $\epsilon > q$) is the one obtained by plugging in λ_3 . Therefore, all eigenvalues are positive if (and only if)

$$\frac{\epsilon}{q} > -2 \cos \left[\frac{\arccos(\mu) - 4\pi}{3} \right] \equiv X(\mu) \quad (10)$$

(which also implies that $\epsilon > q$). Clearly, this is the same as requesting that $\lambda_3 > 0$, since $X(\mu)$ is equal to $(\epsilon - 3\lambda_3)/q$. Using trigonometric identities, $X(\mu)$ can also be expressed as

$$X(\mu) = 2 \cos \left[\frac{\arccos(-\mu)}{3} \right], \quad \text{with } 1 \leq X \leq 2. \quad (11)$$

Its numerical value is also shown in Fig. 1 (dashed line).

Although we have focused on rotational invariants, other combinations of eigenvalues were featured in previous work (Bardeen et al. 1986; Bond & Myers 1996; Sandvik et al. 2007). Quantities built from linear combinations of the eigenvalues tend to all have the same structure: one variable is the trace, which we denote ϵ , and the two other ones are formed from differences with respect to one of the eigenvalues (i.e. they are constructed from the traceless matrix). For example,

$$v_{\pm} = (\lambda_1 - \lambda_3) \pm (\lambda_2 - \lambda_3), \quad (12)$$

which obviously has $v_{\pm} \geq 0$ and $0 \leq v_- \leq v_+$. The shape parameters y and z defined in equation (A.2) of Bardeen et al. (1986) are formed from the differences of the other eigenvalues with respect to λ_2 rather than λ_3 . (The ellipticity and prolateness of Bond & Myers 1996 are conceptually the same: differences with respect to the middle eigenvalue). Our v_{\pm} variables were considered by Sandvik et al. (2007), although they had no real physical justification for their choice. Likewise, one could have worked with differences with respect to λ_1 . We prefer differences with respect to λ_3 because the positivity constraint ($\lambda_3 \geq 0$) is particularly simple in these variables. This is because

$$v_+ = \lambda_1 + \lambda_2 - 2\lambda_3 = \epsilon - 3\lambda_3 = qX(\mu); \quad (13)$$

so, positivity requires $\epsilon \geq 0$ and $0 \leq v_+ \leq \epsilon$, but the v_- range is unchanged ($0 \leq v_- \leq v_+$). As we show below, this makes it easy to describe the joint distribution of (ϵ, v_+, v_-) when all three eigenvalues are positive.

2.2. Protohalo properties

We studied the properties of the protohaloes of haloes identified at $z = 0$ using a spherical overdensity halo finder with a threshold of 319 times the background density, in the Bice and Flora simulations of the SBARBINE suite (Despali et al. 2016). For both simulations, the background cosmology has $\Omega_m = 0.307$, $\Omega_\Lambda = 0.693$, $\sigma_8 = 0.829$ and $h = 0.677$. Each one contains 1024^3 dark matter particles in a periodic box of side $L_{\text{box}} = 125 h^{-1} \text{Mpc}$ (Bice) and $L_{\text{box}} = 2 h^{-1} \text{Gpc}$ (Flora); hence, the corresponding particle masses are $1.55 \times 10^8 h^{-1} M_\odot$ for Bice and $6.35 \times 10^{11} h^{-1} M_\odot$ for Flora.

Our Bice sample contains 5000 randomly selected haloes with masses between 10^{11} and $4 \times 10^{13} h^{-1} M_\odot$. Our Flora sample contains haloes from three separate mass bins, as follows: 2000 randomly selected haloes with masses between 4×10^{13}

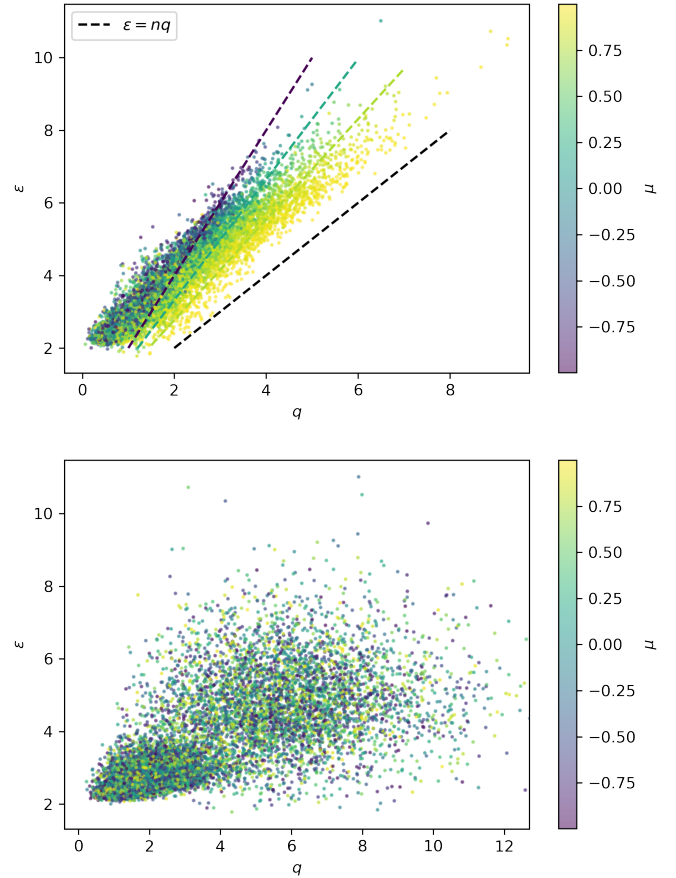


Fig. 2. *Top panel:* Correlation between the three rotational invariants of protohaloes: energy overdensity ϵ , the traceless energy shear q , and μ (see equation 4). The colour displays the value of μ . The dashed lines show $\epsilon = nq$ for $n = \{1, 4/3, 5/3, 2\}$ (from bottom to top), corresponding to the minimum value of ϵ for $\mu = \{1, 0.815, 0.185, -1\}$ whenever the energy shear tensor is positive definite – as shown by equation (10). *Bottom panel:* Expected distribution of $\epsilon - q$ and μ if protohaloes were centred on random locations constrained to have the measured values of ϵ (typically $\epsilon \gtrsim 2$).

and $10^{14} h^{-1} M_\odot$, 2000 randomly selected haloes with masses between 10^{14} and $10^{15} h^{-1} M_\odot$, and all 1378 haloes more massive than $10^{15} h^{-1} M_\odot$. In the following, we will indicate these three samples as Flora-S, Flora-M and Flora-L respectively.

For each halo identified at $z = 0$, we refer to the region occupied by the halo particles in the initial conditions as ‘protohalo’. Following Musso et al. (2024), we measured each protohalo’s centre of mass position and velocity, \mathbf{q}_{cm} and \mathbf{v}_{cm} , by averaging over all its particles. We then estimated its energy shear tensor as

$$\hat{u}_{ij} \equiv -3 \frac{\sum_{n=1}^{N_H} (\mathbf{q}^{(n)} - \mathbf{q}_{\text{cm}})_i (\mathbf{v}^{(n)} - \mathbf{v}_{\text{cm}})_j / fDH}{\sum_{n=1}^{N_H} |\mathbf{q}^{(n)} - \mathbf{q}_{\text{cm}}|^2}, \quad (14)$$

where D is the Λ CDM density perturbation growth factor (at the redshift z of the snapshot), $f = d \ln D / d \ln a$, and n runs over the N_H protohalo particles. (Note that u_{ij} is a ‘dimensionless energy’ because at early times the velocity is proportional to the acceleration, which is the gradient of the gravitational potential: $\nabla\phi \simeq \mathbf{v}/\dot{D} = \mathbf{v}/fDH$). We then diagonalise \hat{u}_{ij} and estimate ϵ , q and μ using Eq. (4).

Figure 2 (top panel) shows a scatter plot of ϵ vs q . The colour indicates the value of μ , measured in the same set of

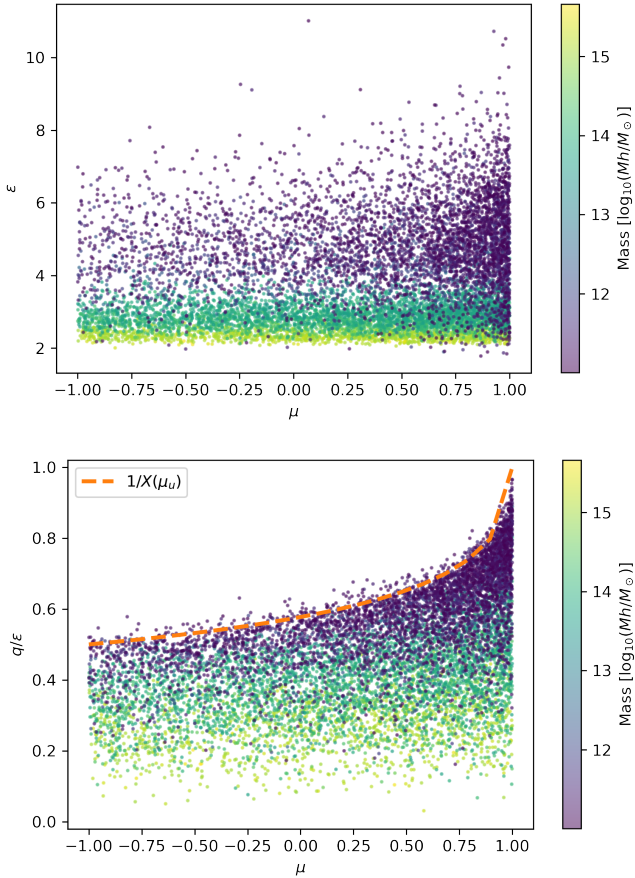


Fig. 3. Correlation of ϵ (top panel) and q/ϵ (bottom) with μ for protohaloes, coloured by $\log_{10}(M/h^{-1} M_{\odot})$. The dashed line in the bottom panel shows the upper limit given by (the inverse of) equation (10).

protohaloes of Musso et al. (2024), where the underlying cosmology and halo identification schemes are described. Indeed, the points here are the same as in their Figure 10; the only difference is that we have coloured each point by its (protohalo’s) value of μ . Clearly, the three invariants are strongly correlated, with higher shear values corresponding to higher ϵ . Moreover, the slope of this nearly linear scaling depends on the third invariant: points with larger μ (coloured in yellow) have shallower ϵ - q slopes, and tend to be lower and to the right compared to points with smaller μ (violet). The latter are confined to the left part of the plot. The dashed lines show the theoretical lower limit for different values of μ : their slopes are $\{1, 4/3, 5/3, 2\}$ from bottom to top, corresponding to $\mu = \{1, 0.815, 0.185, -1\}$.

Before we discuss the dashed lines further, we note that the distribution in the top panel is remarkable because, for generic Gaussian matrices, and thus for tensors measured at generic positions in a Gaussian random field, any variables constructed from the traceless part of the matrix should be uncorrelated with the trace. Therefore, their trace would be a Gaussian variate with zero mean, and their shear an independent χ_5 variate (see discussion following equation 19). The third rotational invariant μ , constructed from the determinant, should also be independent, with a flat distribution between -1 and 1 . Since the distribution in the top panel of Figure 2 clearly has $\epsilon \geq 2$, it is completely different from that for random positions in a Gaussian field.

To test if the distribution of the trace is different, but the distributions of q and μ are not otherwise special, the bottom panel

of Figure 2 shows how the same ϵ - q plot would look if q and μ were independent variables. To generate this plot, we used the measured values of ϵ and σ_{02} , but replaced q and μ with two random variables drawn, respectively, from a χ distribution with 5 degrees of freedom (with the same values of σ_{02}) and a distribution that is uniform between -1 and 1 . This bottom panel is very different from the top one: evidently, the distributions of q and μ are also special.

The correlation shown in the upper panel of Fig. 2 provides some insight into the reason for this behaviour. As a precursor of one of our results, we consider what happens if the energy shear is positive definite, as Musso et al. (2024) showed it must be for protohalo patches. Then, as discussed in Section 2.1, it must be that $\epsilon > q$. This introduces the linear correlation seen in Fig. 2. Secondly, equation (10) must be satisfied, and, since its r.h.s. is a decreasing function of μ , protohaloes with μ closer to -1 are constrained to have larger ϵ/q . This explains why, as one moves away from the lower dashed line in the top panel of Fig. 2 (indicating where $\epsilon = q$), the value of μ tends to decrease. The three steeper dashed lines indicate the lower limit for haloes with ϵ/q larger than $4/3$, $5/3$ and 2 respectively (corresponding to μ of $.185$, $.815$ and 1). Moreover, as shown by equation (8), for a positive definite matrix the inequality $\epsilon > q$ is saturated only if all three eigenvalues simultaneously vanish, which never happens in practice.

Figure 3 (top panel) shows the values of ϵ versus the third rotational invariant μ , coloured by the logarithm of the mass. The yellow and green bands at the bottom correspond to the Flora-L, Flora-M and Flora-S samples, and are quite uniformly distributed between -1 and 1 . Conversely, Bice haloes (lower mass) correspond to the violet dots of the upper band, and show a clear tendency to prefer values closer to 1 . This behaviour can also be understood from equation (10), since for equal values of q protohaloes with $\mu \simeq -1$ must have twice as large values of ϵ as those with $\mu \simeq 1$, and are therefore less likely. Furthermore, except for a few low-mass outliers, in all protohaloes ϵ is larger than roughly 2.2 (we will denote this bottom value ϵ_c in the following), and tends to be higher at smaller masses. The trend of ϵ with mass and q are mutually compatible, since high-mass haloes prefer small values of q and therefore lie in the bottom left corner of Fig. 2.

The bottom panel of Fig. 3 shows the value of the ratio q/ϵ : the upper cut determined by (the inverse of) equation (10) (dashed line in the plot) is clearly visible. This bound affects the low-mass Bice haloes the most. High-mass protohaloes have small q , but nevertheless have (as the top panel shows) non-zero ϵ . Therefore, their q/ϵ is naturally low, and mostly unaffected by the presence of an upper bound.

The fact that the correlation pattern is explained by the $\epsilon > qX(\mu)$ constraint motivates us to look at the correlation structure with the variables $v_+ = qX$ and v_- defined in equation (12). As shown in Figure 4, the linear correlation between ϵ and v_+ is even more prominent, with a scatter that becomes significantly narrower at high values. Points in the graph are coloured by the value of v_-/v_+ in each protohalo: the fact that colours are uniformly distributed in the plot shows that there is no residual correlation of $\epsilon - v_+$ with v_-/v_+ .

2.3. Distribution of rotational invariants of positive-definite matrices

To simplify notation, in this section and in the following ones we set $\sigma_{02} = 1$, unless otherwise specified. The distributions of ϵ and q presented here should therefore be read as those of

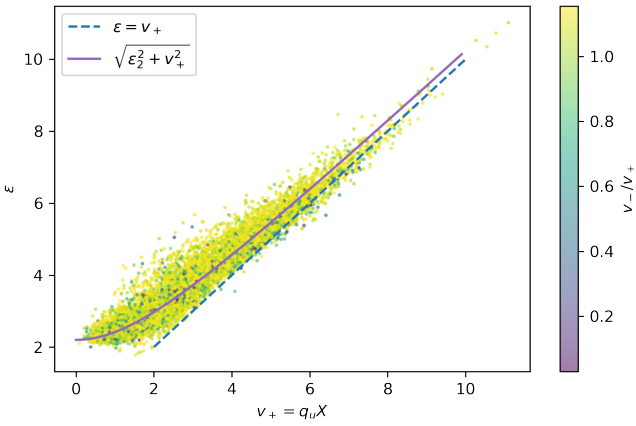


Fig. 4. Correlation of ϵ with v_+ , with colours showing the value of v_-/v_+ . The scatter is significantly narrower than in Figure 2, especially at higher values, and there is no residual correlation signal with v_-/v_+ . Dashed line shows the theoretical lower limit $\epsilon = v_+$, and solid line shows $\sqrt{\epsilon_c^2 + v_+^2}$ with $\epsilon_c = 2.2$.

ϵ/σ_{02} and q/σ_{02} . Since it is defined as a ratio of variables, μ is not affected by this choice.

Doroshkevich (1970) showed that the joint distribution of the three eigenvalues of a 3×3 symmetric Gaussian matrix is given by

$$p(\lambda_1, \lambda_2, \lambda_3) = \frac{3^3 5^3}{8\pi \sqrt{5}} e^{-\epsilon^2/2} e^{-5q^2/2} \Delta(\lambda_1, \lambda_2, \lambda_3), \quad (15)$$

where ϵ and q^2 were defined earlier and

$$\Delta \equiv (\lambda_1 - \lambda_2)(\lambda_1 - \lambda_3)(\lambda_2 - \lambda_3) \quad (16)$$

is the Vandermonde determinant. In terms of the rotational invariants ϵ , q and μ this distribution is

$$p(\lambda_1, \lambda_2, \lambda_3) d\lambda_1 d\lambda_2 d\lambda_3 = \mathcal{N}(\epsilon) \chi(q) \frac{\Theta(1 - \mu^2)}{2} d\epsilon dq d\mu, \quad (17)$$

where

$$\mathcal{N}(\epsilon) d\epsilon = \frac{e^{-\epsilon^2/2}}{\sqrt{2\pi}} d\epsilon \quad (18)$$

and

$$\chi(q) dq = \frac{dq^2}{q^2} \left(\frac{5q^2}{2} \right)^{5/2} \frac{e^{-5q^2/2}}{\Gamma(5/2)}. \quad (19)$$

Therefore, ϵ is Gaussian distributed, $5q^2$ follows a chi-square distribution with five degrees of freedom, and μ is uniform over $[-1, 1]$. In particular, we note that the distributions of ϵ , q and μ are independent; hence, the joint distribution above cannot explain the correlations seen in Figure 2.

This changes when all three eigenvalues have the same sign. If we require $\lambda_3 > 0$, then $q < \epsilon$ and $2\mu > (\epsilon/q)(3 - \epsilon^2/q^2)$. In turn, this requires that $\max[-1, (\epsilon/2q)(3 - \epsilon^2/q^2)] < \mu < 1$. So, integrating over μ returns 1 if $q < \epsilon/2$, but $[(\epsilon/2q)(\epsilon^2/q^2 - 3) + 1]/2$ if $\epsilon/2 < q < \epsilon$. The latter factors as $(\epsilon/q + 2)(\epsilon/q - 1)^2/2$.

Hence, the probability of having all positive eigenvalues is

$$\begin{aligned} p(+) &\equiv p(\lambda_3 \geq 0) \\ &= \int_0^\infty d\epsilon \mathcal{N}(\epsilon) \left[\int_0^{\epsilon/2} dq \chi(q) \right. \\ &\quad \left. + \int_{\epsilon/2}^\epsilon dq \chi(q) \frac{(\epsilon/q + 2)(\epsilon/q - 1)^2}{2} \right] \\ &= \frac{1}{2} - \frac{3 \arctan(\sqrt{5}) + 2\sqrt{5}}{6\pi} \approx \frac{2}{25}. \end{aligned} \quad (20)$$

With this normalisation, the conditional distributions are

$$\begin{aligned} p(\epsilon|+) &= \frac{\mathcal{N}(\epsilon)}{p(+)} \left[\int_0^{\epsilon/2} dq \chi(q) \right. \\ &\quad \left. + \int_{\epsilon/2}^\epsilon dq \chi(q) \frac{(\epsilon/q + 2)(\epsilon/q - 1)^2}{2} \right] \\ &= \frac{\mathcal{N}(\epsilon)}{2p(+)} \left[\operatorname{erf}\left(\frac{\epsilon}{2}\sqrt{\frac{5}{2}}\right) + \operatorname{erf}\left(\epsilon\sqrt{\frac{5}{2}}\right) \right. \\ &\quad \left. - 3\sqrt{5}\epsilon \frac{e^{-5\epsilon^2/8}}{\sqrt{2\pi}} \right], \end{aligned} \quad (21)$$

$$\begin{aligned} p(q|+) &= \frac{\chi(q)}{p(+)} \left[\int_{2q}^\infty d\epsilon \mathcal{N}(\epsilon) \right. \\ &\quad \left. + \int_q^{2q} d\epsilon \mathcal{N}(\epsilon) \frac{(\epsilon/q + 2)(\epsilon/q - 1)^2}{2} \right] \\ &= \frac{\chi(q)}{4p(+)} \left[\operatorname{erfc}(q/\sqrt{2}) + \operatorname{erfc}(q\sqrt{2}) \right. \\ &\quad \left. - \frac{e^{-2q^2}(q^2 + 2) + 2e^{-q^2/2}(q^2 - 1)}{q^3 \sqrt{2\pi}} \right], \end{aligned} \quad (22)$$

$$\begin{aligned} p(\mu|+) &= \frac{1}{2p(+)} \int_0^\infty dq \chi(q) \int_{X(\mu)q}^\infty d\epsilon \mathcal{N}(\epsilon) \\ &= \frac{1}{2\pi p(+)} \left[\operatorname{arccot}\left(\frac{X}{\sqrt{5}}\right) - \frac{\sqrt{5}X(3X^2 + 25)}{3(X^2 + 5)^2} \right], \end{aligned} \quad (23)$$

where $X(\mu)$ was defined in equation (11).

This shows that requiring positive definiteness shifts the distribution of ϵ so that high values are more likely (e.g. its mean is non-zero, close to 1.65); it shifts the distribution of q so that high values are less likely; it induces a correlation between ϵ and q ; and it changes the distribution of μ so that values of unity are more likely. For instance, some of the relevant moments are:

$$\begin{aligned} \langle \epsilon|+ \rangle &\approx 1.665 \approx 5/3, & \langle \epsilon^2|+ \rangle &\approx 3.067, \\ \langle u|+ \rangle &\approx 0.725 \approx 29/40, & \langle u^2|+ \rangle &\approx 0.587. \end{aligned}$$

The cross-correlation between ϵ and q is $C_{\epsilon u} = \langle \epsilon u|+ \rangle \approx 0.068$. The expected slope of the ϵ - q relation is then $C_{\epsilon u}/\operatorname{Var}(u|+) \approx 0.068/0.063 \approx 1$, in agreement with Figure 4.

These distributions of ϵ , q and μ are shown by the dashed lines in Figs. 5 and 6 (the dotted lines correspond to the standard results for Gaussian matrices with unconstrained eigenvalues: normal, chi-5 and uniform). In each figure, the histograms correspond to the measurement of the same quantity in four halo samples (three filled histograms show Flora haloes; open histogram shows the less massive Bice sample). Figure 5 shows that, for high masses (filled histograms), the distribution of ϵ/σ_{02} is strongly mass dependent and is not well described by equation (21). However, for low masses (leftmost histogram) the

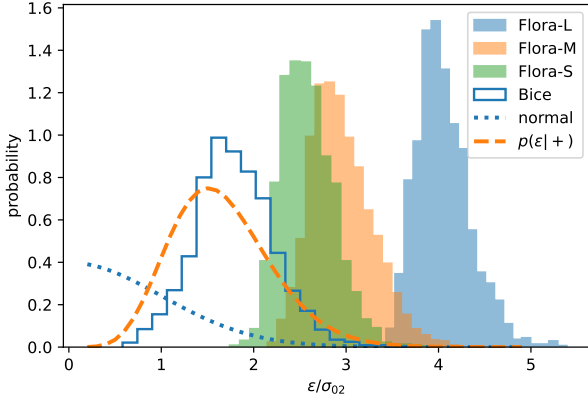


Fig. 5. Distribution of trace ϵ/σ_{02} for a few ranges in halo mass (histograms). The dotted curve shows the expected (zero-mean, unit variance Gaussian) distribution at unconstrained positions; the dashed curve, at positions with $\lambda_3 \geq 0$ (Eq. 21). More massive haloes have larger ϵ/σ_{02} , but the distribution around the mean is approximately independent of mass.

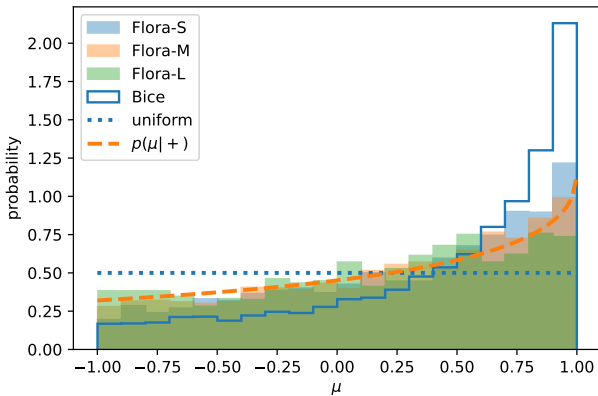
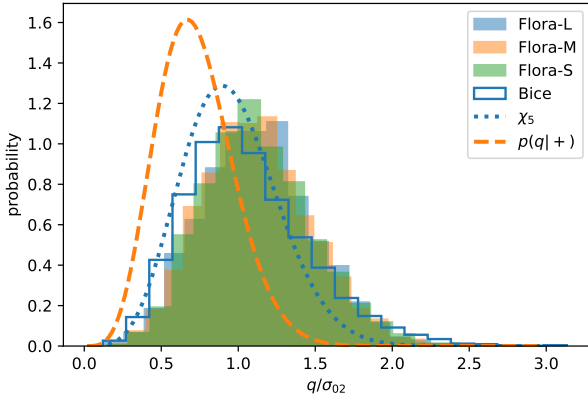


Fig. 6. Distribution of traceless shear q/σ_{02} (top) and μ (bottom). The dotted curves show the expected distributions at unconstrained positions; the dashed ones, at positions with $\lambda_3 \geq 0$ (equations 22 and 23). Whereas q/σ_{02} is approximately the same for all masses, μ is closer to uniform as mass increases.

measurement tends to the predicted (mass-independent) result. This trend can be attributed to the fact that massive protohaloes have $\epsilon \gtrsim 2$ but $q \sim \sigma_{02} \ll 2$; so, they are almost insensitive to the positivity constraint $\epsilon > q$. However, most lower-mass protohaloes have $q > 2$; so, for them $\epsilon > q$ is the dominant constraint, and this constraint explains most of their behaviour.

The top panel of Figure 6 displays the distribution of the traceless shear amplitude q/σ_{02} . In contrast to ϵ/σ_{02} , this shear shows remarkably little mass dependence: the four histograms lie almost on top of each other. Nevertheless, the measured values tend to be higher than predicted by $p(q|+)$ (equation 22), shown by the dashed curve. The bottom panel shows the distribution of μ . In this case, values of μ closer to 1 become more likely, and the trend is more pronounced for less massive haloes. This is a consequence of the constraint $X(\mu) < \epsilon/q$: since small haloes have $\epsilon/q \sim 1$, values of μ closer to 1 are favoured, for which $X(\mu)$ is smaller (see bottom panel of Fig. 3). On the other hand, large haloes typically have $\epsilon/q \gg 1$, which makes the constraint less relevant (since $1 < X(\mu) < 2$). The same can also be inferred from Fig. 2: more and more haloes are excluded as the dashed line becomes steeper, but this does not affect the bottom left corner (where most massive haloes lie).

2.4. Linear combinations of eigenvalues

Figure 4 shows that the $\epsilon - v_+$ relation has smaller scatter than $\epsilon - q$. This motivated a study of the joint distribution of (ϵ, v_+, v_-) , and of how it is affected by the positivity constraint.

Since v_{\pm} are formed from differences of eigenvalues, we expect them to be independent of the trace. Indeed, their joint distribution is (always with the $\sigma_{02} = 1$ convention)

$$p(\epsilon, v_+, v_-) = \frac{15^3}{2^5} \frac{e^{-\epsilon^2/2}}{\sqrt{2\pi}} \frac{e^{-5v_+^2/8 - 15v_-^2/8}}{\sqrt{90\pi}} v_- (v_+^2 - v_-^2), \quad (24)$$

where an unconstrained Gaussian random field has $-\infty \leq \epsilon \leq \infty$, $v_+ \geq 0$ and $0 \leq v_- \leq v_+$. Integrating over v_- yields

$$p(\epsilon, v_+) = \mathcal{N}(\epsilon) \frac{e^{-5v_+^2/8}}{\sqrt{2\pi/5}} \left(e^{-15v_+^2/8} - 1 + 15v_+^2/8 \right). \quad (25)$$

It is straightforward to check that integrating v_+ over the full range ($v_+ \geq 0$) yields $\mathcal{N}(\epsilon)$; and that, recalling that the positivity constraint is simply $0 \leq v_+ \leq \epsilon$ (with the v_- range unchanged), integrating over this restricted range returns $p(\epsilon|+)p(+)$. The distribution of v_+ subject to the constraint that $\lambda_3 \geq 0$ is given by

$$p(v_+|+) = \int_{v_+}^{\infty} d\epsilon \frac{p(\epsilon, v_+)}{p(+)} = \frac{e^{-5v_+^2/8}}{\sqrt{2\pi/5}} \left(e^{-15v_+^2/8} - 1 + 15v_+^2/8 \right) \frac{\text{erfc}(v_+/\sqrt{2})}{2p(+)}. \quad (26)$$

The final error function modifies the shape, effectively reducing the probability of large v_+ .

3. From collapsing to collapsed

The positive-definite constraint is sensible from a dynamical point of view: bound objects should collapse along all three axes. However, Fig. 2 shows a clear lower limit, $\epsilon \gtrsim 2$, that (for small q) is more stringent than the $\epsilon > q$ limit required by positive definiteness (lowest dashed line). This strongly suggests that something in addition to positive definiteness is required to explain the properties of protohaloes (in this case, the patches destined to be haloes by the present time, $z = 0$). Simply requiring $\lambda_3 \geq 0$ does not constrain when the collapse is completed. There may be many objects that are collapsing now but have not yet collapsed completely, and others that collapsed earlier. This raises the question of what sets the time of collapse.

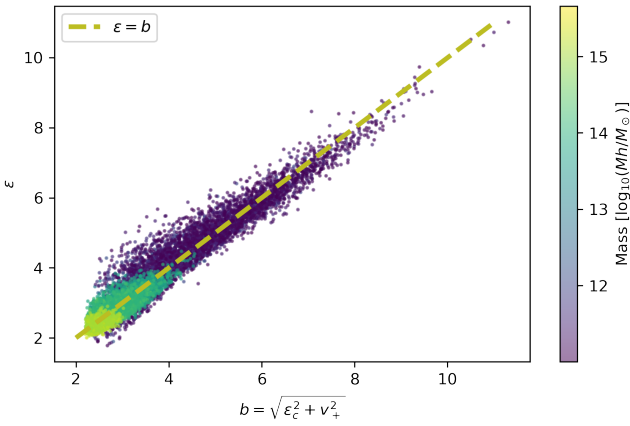


Fig. 7. Scatter plot of ϵ vs $\sqrt{\epsilon_c^2 + v_+^2}$, with $v_+ = qX(\mu)$ and $\epsilon_c = 2.2$, coloured by mass. The residual scatter at fixed v_+ is smaller than the range covered as v_+ changes.

As shown by [Musso & Sheth \(2021\)](#), ϵ controls the evolution of the inertial radius of the halo; for a centrally peaked density profile, ϵ is always larger than mean enclosed matter overdensity, which controls the evolution of the geometric radius (related to the cubic root of the volume). In a simple spherical collapse model, the latter must have an initial overdensity of ~ 1.7 for the required density to be assembled today ([Gunn & Gott 1972](#)). Since the final inertial radius is always smaller than the geometric radius, the threshold on ϵ must be even larger. [Musso & Sheth \(2021\)](#) showed that in large haloes, whose collapse should be fairly anisotropic, ϵ must equal a critical value: $\epsilon_c \approx 2$.

A simple calculation of the distribution of the three rotational invariants ϵ , q and μ subject to the additional constraint that $\epsilon > \epsilon_c$ is given in [Appendix A.1](#). However, the conditional pdf for ϵ resulting from this approach is simply the $\epsilon > \epsilon_c$ tail of equation (21), appropriately normalized. None of our measurements is close to such a sharply cut distribution, and we therefore explored a different path.

In the current context, it is natural to assume that the other rotational invariants also determine the collapse time. Thus, requesting collapse by today enforces a relation

$$C(\epsilon, q, \mu, \epsilon_c) = 0, \quad (27)$$

where C is a function of the three invariants and of a constant ϵ_c that depends on the collapse time. For example, a simple perturbation to the spherical collapse model might be written as $\epsilon \approx \epsilon_c + q$, with $\epsilon_c \approx 2$. Indeed, in models based on density rather than energy it has been proposed that $\delta - \delta_c \propto q$ (e.g. [Sheth et al. 2013](#); [Ludlow et al. 2014](#); [Castorina et al. 2016](#)). This cannot be the full story, though, because [Fig. 2](#) shows that there is substantial scatter in the ϵ - q plane.

As we saw, the scatter is substantially reduced in the ϵ - v_+ plane, where $v_+ = qX$. Moreover, the threshold should be such that $\epsilon > v_+$, in order to guarantee the positivity of all eigenvalues and thus triaxial collapse. Therefore, we considered a threshold that depends on v_+ rather than q . Noting that there appears to be no additional correlation with the other natural variable v_-/v_+ , we supposed that progenitors of haloes identified at the present time satisfy

$$\epsilon = b + y, \quad \text{with} \quad b \equiv \sqrt{\epsilon_c^2 + v_+^2}, \quad (28)$$

where y is some other variable whose variance we assumed to be small, so that the condition $\epsilon > v_+$ is not violated.

For example, a contribution to the scatter at fixed v_+ (hence, to y) would come from differences in the actual virialisation time of haloes identified at the same redshift $z = 0$ ([Borzyszkowski et al. \(2014\)](#)). The details of the virialisation process in fact also depend on whether the initial profile is more or less centrally peaked. This effect is completely independent of the anisotropy of the infall. The solid line in [Figure 4](#) shows $\epsilon = b$ (with $\epsilon_c = 2.2$), and the dashed one shows the lower limit, $\epsilon = v_+$. This plot confirms that the distribution of y is narrower than of b , suggesting that v_+ accounts for most of the measured scatter in ϵ .

To further support our assumption, [Figure 7](#) shows ϵ versus b , coloured by protohalo mass. The range of ϵ values at fixed b (the vertical scatter) is significantly smaller than the range spanned as b changes (the diagonal extension of the plot), especially at low and intermediate masses. This suggests that y can be neglected in first approximation, adopting

$$C = \epsilon - \sqrt{\epsilon_c^2 + v_+^2} = \epsilon - b \approx 0 \quad (29)$$

as the constraint. This assumption reproduces the asymptotic behaviour of the constraint at both large and small v_+ , and automatically guarantees the positivity of all eigenvalues.

Regardless of its exact expression, the threshold depends implicitly on mass through the smoothing scale of the variables. Imposing the constraint therefore introduces a mass dependence in all pdfs, which were so far completely mass independent when expressed in terms of variables normalised by σ_{02} . In fact, if we define the conditional distribution

$$p(\text{variable}|\text{halo of mass } M) \equiv p(\text{variable}|C = 0), \quad (30)$$

where the variable could be any of ϵ/σ_{02} , b/σ_{02} or q/σ_{02} , the constraint introduces an explicit dependence on mass through $\epsilon_c/\sigma_{02}(M)$. Comparing these conditional distributions with measurements in our mass-segregated samples provides further validation of our ansatz on the constraint. To do this comparison properly, however, we need to average over each mass range because each of our samples covers a broad range of masses. We describe this step below.

Before moving on, we note that the most appropriate expression to describe the measurement of a variable in a set of protohaloes of given mass should actually be

$$p(\text{variable}|C = 0, \text{peak}, 1\text{st}), \quad (31)$$

with the conditions including also the peak and first crossing constraint (that is, null gradient, negative-definite Hessian, and threshold not having been crossed on any larger scale). These conditions provide a more accurate description of protohaloes, introducing at the same time new variables that might correlate with the stochasticity in y (which we currently ignore). However, including these effects is more complicated, and the resulting expressions are not always analytical. For these reasons, we continued with the simpler equation (30).

3.1. Constrained q distribution and rescaling of variances

We noted above (histograms of [Fig. 6](#), top panel) that the measured distribution of q/σ_{02} is nearly mass-independent. We need to verify that this remains true for the predicted distribution in the presence of the mass-dependent constraint equation (29). Since $4q^2 = v_+^2 + 3v_-^2$, the conditional distribution for q is

$$p(q|C = 0) = \frac{\langle \delta_D \left(q - \frac{1}{2} \sqrt{v_+^2 + 3v_-^2} \right) \delta_D(C) \rangle}{\langle \delta_D(C) \rangle}, \quad (32)$$

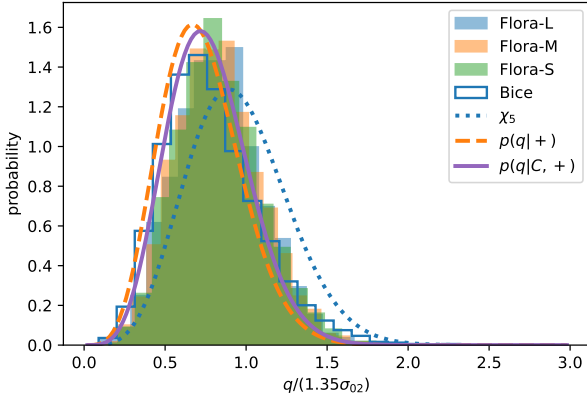


Fig. 8. Distribution of $q/(1.35\sigma_{02})$, the rescaling that provides the best match to equation (34) (purple solid line). Histograms show the measurements; dotted, dashed and solid curves show the distribution of random positions, those constrained to be positive definite, and those constrained to also have $\epsilon = b$.

where $\langle \dots \rangle$ means that we averaged using equation (24) (in the numerator) or equation (25) (in the denominator). Note that the positivity condition is automatically implied by enforcing $C = 0$.

The denominator of the expression above evaluates to

$$\langle \delta_D(C) \rangle = \frac{e^{-\epsilon_c^2/2}}{36\sqrt{\pi}} \sqrt{5(29 - 6\sqrt{6})}; \quad (33)$$

thus, replacing $\epsilon_c \rightarrow \epsilon_c/\sigma_{02}(M)$ to reintroduce σ_{02} , it is exponentially mass dependent. However, the exponential cancels out with a similar term in the numerator, and we get

$$\begin{aligned} p(q|C=0) &= \frac{15^2 \sqrt{2} q e^{-5q^2/2}}{\sqrt{(29 - 6\sqrt{6})}} \int_q^{2q} dv \frac{e^{-v^2/2}}{\sqrt{2\pi}} (v^2 - q^2) \\ &= \frac{15^2 q e^{-5q^2/2}}{\sqrt{\pi(29 - 6\sqrt{6})}} \left[q \left(e^{-q^2/2} - 2e^{-2q^2} \right) \right. \\ &\quad \left. + \sqrt{\frac{\pi}{2}} (q^2 - 1) \left(\operatorname{erf}\left(\frac{q}{\sqrt{2}}\right) - \operatorname{erf}(\sqrt{2}q) \right) \right], \quad (34) \end{aligned}$$

where there is no dependence on ϵ_c , and thus no mass dependence left. This expression is numerically quite close to equation (22); in particular, it is closer to the dashed curve in Fig. 6 than to the histograms there. In other words, using a better model for the threshold does not fix the disagreement with the measured distribution of q for protohaloes.

We should not expect to find perfect agreement, however, since protohalo patches are not spherical, whereas our analysis implicitly assumed they are. This for instance changes the prediction of the variance. To account at least qualitatively for this, we rescaled σ_{02} (the standard deviation of ϵ in spheres containing a given mass) by an ad hoc factor whose value we motivate elsewhere (see Padmanabhan & Subramanian 1993; Betancort-Rijo & López-Corredoira 2002; Lam & Sheth 2008, for earlier work). Here, we simply note that this factor should be greater than unity, to compensate for the fact that ϵ in the protohalo shapes is always larger than in the corresponding Lagrangian sphere (Musso & Sheth 2023). We also assumed for simplicity that this factor is the same for all masses. To fix its amplitude, we required that the distribution of q/σ_ϵ match equation (34) closely. Fig. 8 shows that setting $\sigma_\epsilon = 1.35\sigma_{02}$ provides a reasonably good agreement: the rescaled histograms are

closer to the solid curve than to the dotted one. Presumably, some of the small differences that remain are due to approximating equation (31) with equation (30), and ignoring the stochasticity that allowed us to approximate equation (28) with equation (29).

3.2. Testing the threshold ansatz

We tested if equation (29) is a useful parametrisation of the threshold for collapse by comparing the measured distributions of ϵ/σ_{02} and b with the predicted conditional distribution. The latter can be expressed as

$$p(b|C=0) = \frac{\int_0^b dv_+ \delta_D \left(b - \sqrt{\epsilon_c^2 + v_+^2} \right) p(\epsilon = b, v_+)}{\int_0^\infty d\epsilon \int_0^\epsilon dv_+ \delta_D \left(\epsilon - \sqrt{\epsilon_c^2 + v_+^2} \right) p(\epsilon, v_+)}; \quad (35)$$

Explicit calculation, using $p(\epsilon, v_+)$ from equation (25), yields

$$p(b|C=0) = \frac{18b e^{-9v_+^2/8} e^{-15v_+^2/8} - 1 + 15v_+^2/8}{\sqrt{\pi v_+^2} \sqrt{29 - 6\sqrt{6}}} \Big|_{v_+ = b^2 - \epsilon_c^2}, \quad (36)$$

where $2b db = 2v_+ dv_+$ is the reason why there is a b/v_+ factor in the expression above. Of course, this assumes that $b \geq \epsilon_c$, and $p(b|C=0) = 0$ otherwise.

This distribution explicitly depends on ϵ_c , that is, after reintroducing σ_{02} , on $\epsilon_c/\sigma_{02}(M)$, so we must be careful when comparing it with measurements that span a range of masses. (This was not a concern for q/σ_{02} , for which there was no predicted mass dependence). To make a fair comparison, we averaged equation (36) over the same set of masses that contributed to the measurement. This average is

$$p(b|[M_1, M_2]) = \frac{\int_{M_1}^{M_2} dM n(M) p(b|C=0)}{\int_{M_1}^{M_2} dM n(M)}, \quad (37)$$

where $n(M)$ is the mass function. Ideally, we would like to predict $n(M)$, rather than using the measured distribution. While a more careful derivation is the subject of work in progress, the plots that follow were made using the actual distribution of masses measured in each sample. We provide a simple estimate for $n(M)$ motivated by an approach à la Press–Schechter in Appendix A.2.

With this in mind, we proceeded to a quantitative comparison. The filled histograms of Fig. 9 show the distribution of b in three of our protohalo samples, and the dashed step lines the measured distribution of ϵ (i.e. the same histograms as in Fig. 5). The similarity of the two sets of histograms proves that setting $\epsilon^2 = \epsilon_c^2 + v_+^2$ accounts for most of the scatter and is therefore a very good approximation for the threshold for halo formation.

The measured histograms are shifted to higher values of b/σ_{02} (compared to the theory curves). In our discussion of q , we noted that rescaling $\sigma_{02} \rightarrow 1.35\sigma_{02}$ worked quite well; hence, it is natural to ask if this also works well here. Fig. 10 shows that it does. As further confirmation of the accuracy of this rescaling, Fig. 11 compares the distribution of $b/1.35$ measured in narrow mass bins directly with equation (36). We chose three mass intervals of $(1-2) \times 10^{11}$, $(4-5) \times 10^{13}$ and $(1-1.25) \times 10^{15} h^{-1} M_\odot$, corresponding to the lower end of Bice, Flora-S and Flora-L respectively. The bins are sufficiently narrow that no averaging over σ_{02} values should be needed. Therefore, for the theory curves we use as ω_c the mean of $\epsilon_c/(1.35\sigma_{02})$ in each narrow bin.

The remarkable agreement with the filled histograms confirms that simply requiring the positivity of the eigenvalues, and

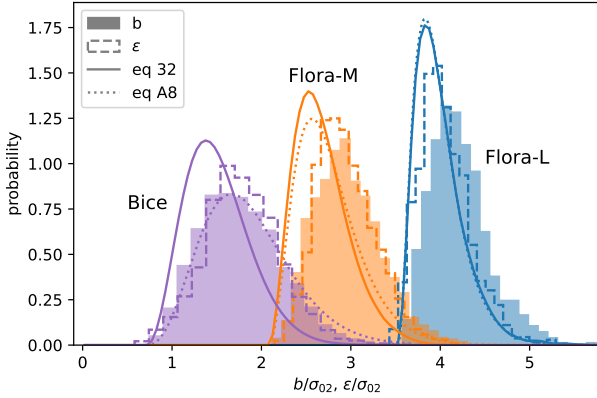


Fig. 9. Distribution of $b = \sqrt{\epsilon_c^2 + v_+^2} / \sigma_{02}$ in the three mass bins (filled histograms). The dashed lines show the distribution of ϵ / σ_{02} (from Figure 5), the solid lines show equation (36) averaged over the measured distribution of σ_{02} values, and the dotted lines show equation (A.8).

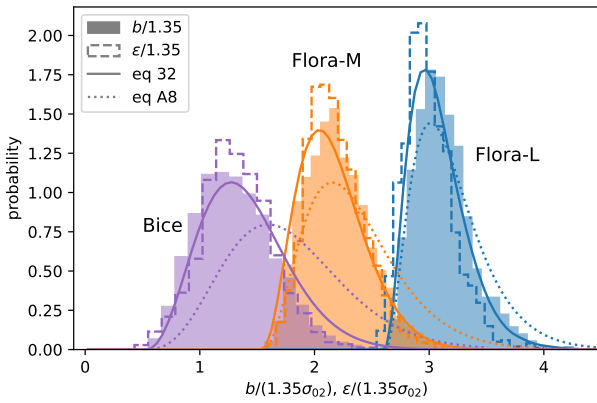


Fig. 10. Weighted average of equation (36) (solid curves) with $\epsilon_c \rightarrow \epsilon_c / (1.35\sigma_{02})$, as described in the main text, providing an excellent description of the rescaled histograms of $b / (1.35\sigma_{02})$ for all samples. The dotted curves show equation (A.8).

imposing the approximate constraint $\epsilon^2 = \epsilon_c^2 + v_+^2$, gets very close to describing the correct distribution of v_+ (and thus of b) in protohaloes. In turn, b is a rather good approximation to the value of ϵ , and therefore to the collapse threshold of protohaloes. While a more careful calculation would require that we also impose the excursion set and peak constraints (see equation 31 and related discussion), we believe that the considerably simpler calculation discussed here (equations 29 and 30) captures most of the interesting physics.

4. Discussion and conclusions

We explored the statistical implications of two facts of dynamical origin that make protohaloes – the initial patches that are destined to become haloes – special locations: they must collapse along three axes, and they have to do so by today. Expressed in mathematical terms, these conditions are:

- the energy shear u_{ij} is positive definite, so that all three axes of the protohalo collapse (Musso et al. 2024);
- ϵ , the trace of u_{ij} , must exceed a critical value for the collapse to complete by the present day (Musso & Sheth 2021).

Both conditions are interesting in their own right.

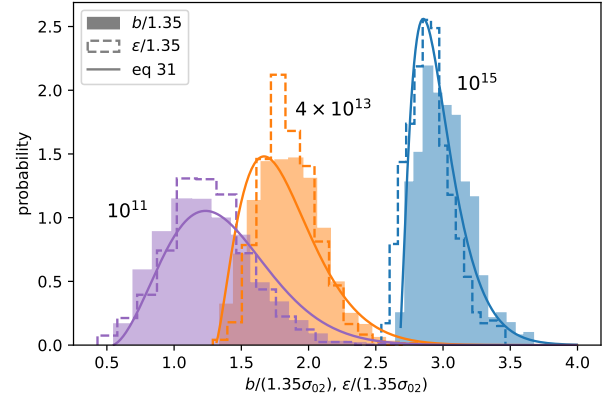


Fig. 11. Distribution of $b / (1.35\sigma_{02})$ (filled histograms) and $\epsilon / (1.35\sigma_{02})$ (empty dashed) in three narrow mass bins of mass 10^{11} , 4×10^{13} and $10^{15} h^{-1} M_\odot$ (from left to right). The solid curves show equation (36).

The energy shear tensor u_{ij} (equation 1) measured at random positions in a Gaussian random field is a Gaussian random matrix, and its three rotational invariants (equation 4) are therefore uncorrelated with one another. However, in protohaloes they are strongly correlated (Fig. 2). We showed that the measured correlation pattern – in particular, the limiting values of some of these correlations (Figs. 2 and 3) – directly follows from the positive definiteness of the energy shear, that is from condition (a). We provided analytic expressions for the distributions of the three invariants, and their correlations, subject to these conditions (equations 21–23, and A.3–A.5 if one also requests $\epsilon > \epsilon_c$).

Previous work showed that the collapse threshold in ϵ is stochastic, with an absolute lower bound that we denoted ϵ_c . Fig. 2 shows that, for the protohaloes of haloes identified at $z = 0$, $\epsilon_c \approx 2$. Although condition (a) cannot explain the lower bound, the positivity requirement means that the excursion above this limit strongly correlates with the other rotational invariants (built out of differences of the eigenvalues). We verified that a parametrization of the stochasticity based on v_+ (equation 12) appears to be more efficient – that is, it has significantly less residual scatter – than one based on the rotational invariants q and μ (compare Figs. 2 and 4). In particular, ϵ in protohaloes is very close to $b = \sqrt{\epsilon_c^2 + v_+^2}$ (Fig. 7). Parametrising the threshold with this expression accounts for most of its stochasticity.

We then computed the conditional distribution of q given that ϵ equals the threshold. The result is rather insensitive to the constraint, and it matches the measured distributions well provided that one rescales σ_{02} (the standard deviation of ϵ in spherical volumes) by a factor of 1.35 (Fig. 8). We argued that this rescaling can be attributed to the fact that protohaloes are not spherical. Computing it from first principles is the object of current work. The conditional distribution of b (equation 36), on the other hand, is very sensitive to the constraint, which makes it strongly mass dependent. However, the same variance rescaling brings the measurements of this quantity into good agreement with our (mass-dependent) predictions.

To summarise, we began our study with two main goals. One was the question of whether μ , the third invariant of the energy shear tensor, matters for protohaloes. Fig. 2 shows that it does, which has therefore potential implications for assembly bias. Our subsequent analysis suggests that a more efficient parametrization of the physics of collapse uses v_+ , a non-linear combination of the second and third invariants (q and μ) that is actually linear in the eigenvalues themselves (equation 13).

The second goal was to determine if the correlations between the invariants of the energy shear measured at protohalo positions (at random positions they are independent) can be understood as simply arising from the positivity of the eigenvalues. Here too the answer is yes, but with some subtleties. Our analysis suggests that triaxial collapse requires $\lambda_3 \geq 0$, but collapse by today means $\epsilon \geq 2$. For very massive protohaloes, the dominant condition is the latter: in this regime, $\epsilon \sim 2$ with very little residual dependence on other rotational invariants. For smaller objects, however, which typically have $\epsilon \gg 2$, much of the correlation between ϵ and q or v_+ emerges from the constrained statistics, rather than being an additional physical requirement. The positivity condition is much simpler in the basis (ϵ, v_+, v_-) than in (ϵ, q, μ) , which explains why parametrising the stochasticity of the threshold with v_+ is more efficient.

Our predictions are completely analytic and reasonably simple. This contrasts with many results on protohalo statistics, typically based on peak theory and first passage probabilities, which require rather involved calculations. These complications are indeed necessary when computing number densities, masses and correlation functions. However, a byproduct of our analysis is that they are not always required to predict the probability of secondary quantities in protohaloes. For these, conditioning on a sufficiently good model of the threshold is enough.

These results bring us closer to a first-principles derivation of the abundance of dark matter haloes and their properties. Whereas it is quite well established that haloes originate from sufficiently high peaks of the initial field smoothed on the corresponding mass scale, how tall these peaks should exactly be is a point under active discussion. Previous research has shown that the threshold is not the same for all haloes: it contains significant scatter. Accounting for this scatter is crucial to correctly predict when each halo virialised. Moreover, there is considerable interest in the role of additional variables on halo formation and clustering (sometimes called ‘assembly’ or ‘secondary’ bias), and which observable properties of the final halo they determine. We anticipate that some of the distributions of these secondary quantities, which we derived in Section 3, may be helpful in modelling these effects.

Three specific issues that still need to be addressed are:

- (i) the effect of non-sphericity of protohalo shapes on their statistics;
- (ii) the effect on the statistics of accounting for the fact that protohalo positions define a point process; and
- (iii) the exact relation between the value of ϵ and the time of virialisation.

About the first point, our work suggests that the main effect of initial non-sphericity is to rescale all variances. In a follow-up to Musso & Sheth (2023), who required protohalo shapes to maximise ϵ , we are in the process of showing why this rescaling takes place. In addition, now that we have a good model for the critical threshold (equation 28), we believe we are in a position to apply the technology of excursion set energy peaks theory (Musso & Sheth 2021) to address point (ii). This will allow us to quantify assembly or secondary bias effects as well (Sheth & Tormen 2004; Dalal et al. 2008; Musso & Sheth 2014; Lazeyras et al. 2017; Matsubara 2019; Lazeyras et al. 2023). It might also help us to improve models of shape bias for intrinsic alignments (Blazek et al. 2015; Vlah et al. 2020; Lee & Moon 2024; Maion et al. 2025)

As for point (iii), we note that a halo identified at $z = 0$ may have virialised at an earlier time: halo finders do not usually make this distinction. Since in first approximation the evolution of the inertial radius follows spherical collapse, where

the final time depends on the initial energy, the uncertainty in the virialisation time will add to the stochasticity in the value of ϵ (similarly to Borzyszkowski et al. 2014). Much of this stochasticity, as shown in this work, is described by v_+ : when v_+ is large then so is ϵ , implying early virialisation. Moreover, $\epsilon \sim v_+$ means that $\lambda_3 \sim 0$: these objects represent the transition between a protohalo and a protofilament (Hanami 2001; Cadiou et al. 2020; Salvador-Solé & Manrique 2021), and can be associated with haloes that stopped accreting before the time of identification due to tidal stripping (as found in simulations by Borzyszkowski et al. 2017). There is however also a residual scatter in ϵ at fixed v_+ , which might be related to how virialisation depends on the slope of the initial profile. We hope to quantify this effect in future work.

To obtain a more reliable estimate of the virialisation time in an anisotropic environment (large v_+), we may consider perturbative corrections to spherical collapse. These corrections naturally scale like q^2/ϵ^2 . Our analysis suggests that this ratio cannot exceed unity, which helps us to keep the perturbative treatment under control. Physically, this can be interpreted as reflecting the fact that the evolution cannot deviate too much from spherical collapse, even for arbitrary shapes, simply because the protohalo patch must end up in a very small region in order to create a high enough density. This is what makes protohaloes special locations.

Finally, our goal is to predict where haloes form in the *initial* field. Recent machine-learning based work aims to predict where they occur in the *evolved* field (Pandey et al. 2024). We hope that our work will be useful to train these algorithms to make predictions from the initial field.

Data availability

The simulation data and the post-processing quantities used in this work can be shared on reasonable request to the authors.

Acknowledgements. We are grateful to the anonymous referee for a helpful report. Thanks to G. Despali for sharing the data from the SBARBINE simulations, and to the organisers and participants of the 2023 KITP Cosmic Web workshop. Our visit to KITP was supported by the National Science Foundation under PHY-1748958. We are grateful to the EAFIR, Kigali and the ICTP, Trieste for their hospitality in summer 2024, when this work was completed. Marcello Musso acknowledges support from grants PID2021-122938NB-I00, PID2024-158938NB-I00 and CNS2024-154286 funded by MICIU/AEI/10.13039/501100011033 and by “ERDF A way of making Europe”, and SA097P24 funded by Junta de Castilla y León.

References

- Bardeen, J. M., Bond, J. R., Kaiser, N., & Szalay, A. S. 1986, *ApJ*, 304, 15
- Betancort-Rijo, J., & López-Corredoira, M. 2002, *ApJ*, 566, 623
- Blazek, J., Vlah, Z., & Seljak, U. 2015, *JCAP*, 08, 015
- Bond, J. R., & Myers, S. T. 1996, *ApJS*, 103, 1
- Borzyszkowski, M., Ludlow, A. D., & Porciani, C. 2014, *MNRAS*, 445, 4124
- Borzyszkowski, M., Porciani, C., Romano-Diaz, E., & Garaldi, E. 2017, *MNRAS*, 469, 594
- Cadiou, C., Pichon, C., Codis, S., et al. 2020, *MNRAS*, 496, 4787
- Castorina, E., Paranjape, A., Hahn, O., & Sheth, R. K. 2016, ArXiv e-prints [arXiv:1611.03619]
- Dalal, N., White, M., Bond, J. R., & Shirokov, A. 2008, *ApJ*, 687, 12
- Desjacques, V. 2013, *Phys. Rev. D*, 87, 043505
- Despali, G., Tormen, G., & Sheth, R. K. 2013, *MNRAS*, 431, 1143
- Despali, G., Giocoli, C., Angulo, R. E., et al. 2016, *MNRAS*, 456, 2486
- Doroshkevich, A. G. 1970, *Astrophysics*, 6, 320
- Gunn, J. E., & Gott, J. R. 1972, *ApJ*, 176, 1
- Hanami, H. 2001, *MNRAS*, 327, 721
- Lam, T. Y., & Sheth, R. K. 2008, *MNRAS*, 386, 407
- Lazeyras, T., & Schmidt, F. 2018, *JCAP*, 2018, 008

- Lazeyras, T., Musso, M., & Desjacques, V. 2016, [Phys. Rev. D](#), **93**, 063007
- Lazeyras, T., Musso, M., & Schmidt, F. 2017, [JCAP](#), **2017**, 059
- Lazeyras, T., Barreira, A., Schmidt, F., & Desjacques, V. 2023, [JCAP](#), **2023**, 023
- Lee, J., & Moon, J.-S. 2024, [JCAP](#), **10**, 102
- Ludlow, A. D., Borzyszkowski, M., & Porciani, C. 2014, [MNRAS](#), **445**, 4110
- Maion, F., Sticker, J., & Angulo, R. E. 2025, [A&A](#), **699**, A271
- Matsubara, T. 2019, [Phys. Rev. D](#), **100**, 083504
- Modi, C., Castorina, E., & Seljak, U. 2017, [MNRAS](#), **472**, 3959
- Musso, M., & Sheth, R. K. 2014, [MNRAS](#), **438**, 2683
- Musso, M., & Sheth, R. K. 2021, [MNRAS](#), **508**, 3634
- Musso, M., & Sheth, R. K. 2023, [MNRAS](#), **523**, L4
- Musso, M., Despali, G., & Sheth, R. K. 2024, [MNRAS](#), **523**, L4
- Padmanabhan, T., & Subramanian, K. 1993, [ApJ](#), **410**, 482
- Pandey, S., Lanusse, F., Modi, C., & Wandelt, B. D. 2024, [41st International Conference on Machine Learning](#)
- Press, W. H., & Schechter, P. 1974, [ApJ](#), **187**, 425
- Salvador-Solé, E., & Manrique, A. 2021, [ApJ](#), **914**, 141
- Sandvik, H. B., Möller, O., Lee, J., & White, S. D. M. 2007, [MNRAS](#), **377**, 234
- Sheth, R. K., & Tormen, G. 2002, [MNRAS](#), **329**, 61
- Sheth, R. K., & Tormen, G. 2004, [MNRAS](#), **350**, 1385
- Sheth, R. K., Mo, H., & Tormen, G. 2001, [MNRAS](#), **323**, 1
- Sheth, R. K., Chan, K. C., & Scoccimarro, R. 2013, [Phys. Rev. D](#), **87**, 083002
- van de Weygaert, R., & Bertschinger, E. 1996, [MNRAS](#), **281**, 84
- Vlah, Z., Chisari, N. E., & Schmidt, F. 2020, [JCAP](#), **01**, 025

Appendix A: Statistics for a range of masses

We discuss a simple model for the statistics when the sample includes objects more massive than a given threshold. The approach is similar to that of [Press & Schechter \(1974\)](#), which is known to not be very accurate – in particular, it does not account for the fact that haloes define a point process. Nevertheless, we believe it provides some insight.

A.1. Simple threshold: $\epsilon = \epsilon_c$

Integrating over masses above a threshold is like integrating σ_{02} from 0 to an upper limit. Since the statistics depend on ϵ_c/σ_{02} , this is like integrating ϵ_c from a lower limit to infinity.

Hence, as a warm up, consider simply requiring $\epsilon \geq \epsilon_c$. To quantify how the distribution of ϵ , q and μ are modified, we first define the new normalisation

$$p(+\epsilon_c) \equiv p(\lambda_3 \leq 0, \epsilon > \epsilon_c) = p(+)\int_{\epsilon_c}^{\infty} d\epsilon p(\epsilon|+). \quad (\text{A.1})$$

Although no analytical expression for $p(+\epsilon_c)$ exists, a simple and reasonably accurate approximation is

$$p(+\epsilon_c) \simeq \frac{1}{4} \operatorname{erfc}\left(\frac{\epsilon_c}{\sqrt{2}\sigma_{02}}\right) \left[1 + \operatorname{erf}\left(\frac{\epsilon_c}{\sqrt{2}\sigma_{02}}\right)\right] - \frac{\sqrt{5}}{3\pi} e^{-9\epsilon_c^2/8\sigma_{02}^2} + \frac{\pi - 2 \arctan(\sqrt{5})}{4\pi} e^{-3\epsilon_c^2/2\sigma_{02}^2}, \quad (\text{A.2})$$

which captures the correct asymptotic behaviour for small and large σ_{02} (large and small masses) and deviates from the numerical solution by less than 1.8% over the entire positive axis.

Equations (21), (22) and (23) are changed to

$$p(\epsilon|+\epsilon_c) = \Theta(\epsilon - \epsilon_c) p(\epsilon|+) \frac{p(+)}{p(+\epsilon_c)}, \quad (\text{A.3})$$

$$p(q|+\epsilon_c) = \frac{\chi(q)}{p(+\epsilon_c)} \left[\int_{q_{\min 2}}^{\infty} d\epsilon p(\epsilon) + \int_{q_{\min 1}}^{q_{\min 2}} d\epsilon p(\epsilon) \frac{(\epsilon/2q)(\epsilon^2/q^2 - 3) + 1}{2} \right] = \frac{\chi(q)}{4p(+\epsilon_c)} \left[\operatorname{erfc}\left(\frac{q_{\min 1}}{\sqrt{2}}\right) + \operatorname{erfc}\left(\frac{q_{\min 2}}{\sqrt{2}}\right) - \frac{e^{-q_{\min 2}^2/2} (q_{\min 2}^2 + 2 - 3q^2)}{\sqrt{2\pi} q^3} + \frac{e^{-q_{\min 1}^2/2} (q_{\min 1}^2 + 2 - 3q^2)}{\sqrt{2\pi} q^3} \right] \quad (\text{A.4})$$

where $q_{\min n} \equiv \max(\epsilon_c/\sigma_{02}, nq)$, and

$$p(\mu|+\epsilon_c) = \frac{1}{2p(+\epsilon_c)} \int_{\epsilon_c}^{\infty} d\epsilon p(\epsilon) \int_0^{\epsilon/X} dq \chi(q) = \int_{\epsilon_c}^{\infty} d\epsilon \frac{p(\epsilon)}{2p(+\epsilon_c)} \left[\operatorname{erf}\left(\sqrt{\frac{5}{2}} \frac{\epsilon}{X}\right) - \sqrt{\frac{10}{\pi}} \epsilon \frac{5\epsilon^2 + 3X^2}{3X^3} e^{-5\epsilon^2/2X^2} \right]. \quad (\text{A.5})$$

The last integral does not have an analytical expression, but an approximation accurate to better than 1.5% is

$$p(\mu|+\epsilon_c) \simeq \frac{1}{p(+\epsilon_c)} \left[\frac{1}{8} \left(\operatorname{erf}\left(\frac{\omega_c}{\sqrt{2}}\right) + 1 \right) \operatorname{erfc}\left(\frac{\omega_c}{\sqrt{2}}\right) + \left(\frac{\arctan(\sqrt{5}/X)}{2\pi} - \frac{1}{8} \right) e^{-(X^2+5)\omega_c^2/4X} \right]. \quad (\text{A.6})$$

Massive objects correspond to the $\epsilon_c/\sigma_{02} \gg 1$ limit, for which $p(+\epsilon_c) \simeq \frac{1}{2} \operatorname{erfc}(\epsilon_c/\sqrt{2}\sigma_{02})$ and $q_{\min 1} = q_{\min 2} = \epsilon_c/\sigma_{02}$. Therefore, only the first term of equation (22) contributes and tends to $\chi(q)$ of Eq. (19) (dotted curve in Fig. 6). In contrast, at small masses the limits of the integrals in Eq. (22) are modified only when $q \leq 2$ (i.e. not significantly). As a result, lower mass haloes are better described by Eq. (22) (dashed curve in Fig. 6).

These limiting cases can also be understood by looking at Fig. 2 which shows the two cuts explicitly: the dashed line showing $\epsilon = q$ and the obvious floor at $\epsilon = 2$. Most lower mass haloes tend to have large q , so they are less affected by the $\epsilon > 2$ floor: the $\epsilon > q$ cut matters much more. However, massive haloes tend to have small q , so the $\epsilon > 2$ floor is the only relevant cut. In a Gaussian field, the distribution of q is uncorrelated with ϵ . This is modified by the $\epsilon = q$ cut, but, since this cut is irrelevant for the massive haloes, their distribution is not given by Eq. (22); instead, it should be much closer to Eq. (19). Fig. 6 shows that the distribution of massive haloes is indeed shifted towards larger values, as expected.

By this argument, one would expect the distribution of μ for massive haloes to be closer to that for unconstrained positions (uniform on $[-1, 1]$); this is in qualitative agreement with the trend in the bottom panel of Fig. 6, where the largest mass objects (green) are closest to being uniformly distributed.

A.2. More realistic threshold: $\epsilon = \sqrt{\epsilon_c^2 + v_+^2}$

Now suppose the threshold is given by equation (28) of the main text (and the extra variable y is ignored). Integrating over all masses above some minimum means we now want $\sqrt{\epsilon^2 - v_+^2}/\sigma_{02} > \omega_c$, or equivalently $v_+/\sigma_{02} < \sqrt{\omega^2 - \omega_c^2}$. The resulting distribution is then

$$p(b|C > 0) \propto \int_0^{\sqrt{b^2 - \omega_c^2}} dv_+ p(\epsilon = b, v_+), \quad (\text{A.7})$$

where the integral is analytical and is easily seen to be equal to $e^{-\omega_c^2/2} p(\sqrt{b^2 - \omega_c^2}|+) p(+)$. Normalising to unity, we get

$$p(b|C > 0) = \frac{p(\epsilon = \sqrt{b^2 - \omega_c^2}|+)}{N(\omega_c)}, \quad (\text{A.8})$$

where $N(\omega_c)$ is the integral of the numerator over $b > \omega_c$. Changing variables from b to $\sqrt{b^2 - \omega_c^2}$, the normalisation becomes

$$N(\omega_c) = \int_0^{+\infty} d\epsilon \frac{\epsilon}{\sqrt{\epsilon^2 + \omega_c^2}} p(\epsilon|+); \quad (\text{A.9})$$

this integral has no known analytical solution, but can be approximated by Taylor-expanding the fraction $\epsilon/\sqrt{\epsilon^2 + \omega_c^2}$ around the conditional mean $\langle \epsilon|+ \rangle \simeq 1.663$. At second order, this gives

$$N(\omega_c) \simeq \frac{\langle \epsilon|+ \rangle}{\sqrt{\langle \epsilon|+ \rangle^2 + \omega_c^2}} \left[1 - \frac{3\operatorname{Var}(\epsilon|+)}{(\langle \epsilon|+ \rangle^2 + \omega_c^2)^2} \frac{\omega_c^2}{2} \right], \quad (\text{A.10})$$

where $\text{Var}(\epsilon|+) \simeq 0.3010$. This approximation is asymptotically correct for both large and small ω_c (in which limits one gets $\langle \epsilon|+ \rangle / \omega_c$ and 1 respectively), and numerically accurate to less than 0.4% everywhere. The dotted curves in Fig. 9 show equation (A.8). While the agreement with the measurements is not compelling, as we noted above, this is only a crude estimate. A better estimate would include the peak and first-crossing constraints, and the corresponding Jacobians: the Hessian determinant and the slope of the excursion set (Musso & Sheth 2021). This is the subject of ongoing work.