Symmetry in stochasticity: Random walk models of large-scale structure

RAVI K SHETH[1,2]

[1]Center for Particle Cosmology, University of Pennsylvania, Philadelphia, PA 19104, USA
[2]The Abdus Salam Center for Theoretical Physics, Strada Costiera 11, 34151 Trieste, Italy
E-mail: shethrk@physics.upenn.edu

Abstract. This paper describes the insights gained from the excursion set approach, in which various questions about the phenomenology of large-scale structure formation can be mapped to problems associated with the first crossing distribution of appropriately defined barriers by random walks. Much of this is summarized in R K Sheth, *AIP Conf. Proc.* **1132**, 158 (2009). So only a summary is given here, and instead a few new excursion set related ideas and results which are not published elsewhere are presented. One is a generalization of the formation time distribution to the case in which formation corresponds to the time when half the mass was first assembled in pieces, each of which was at least $1/n$ times the final mass, and where $n \geq 2$; another is an analysis of the first crossing distribution of the Ornstein–Uhlenbeck process. The first derives from the mirror-image symmetry argument for random walks which Chandrasekhar described so elegantly in 1943; the second corrects a misuse of this argument. Finally, some discussion of the correlated steps and correlated walks assumptions associated with the excursion set approach, and the relation between these and peaks theory are also included. These are problems in which Chandra's mirror-image symmetry is broken.

Keywords. Galaxies – formation; large-scale structure; cosmology.

PACS Nos 98.65.Cw; 98.65.Dx

1. The excursion set approach

Virialized objects had to fight the expansion of the Universe for their formation. This fight is more easily won if they had a head start – if they grew from large initial overdensities. Thus, given a model for gravity, the abundance of virialized objects contains information about the initial fluctuation field, and about the subsequent expansion history of the Universe. This underlying philosophy is sketched in figure 1. The excursion set approach [1–4] was developed to describe how this information is encoded in the abundance and clustering of nonlinear structures present at later times, and in their formation histories.

The key to this approach is the assumption that the nonlinear field has some memory of the initial conditions. To see why, suppose we choose a random particle in the initial density fluctuation field, and imagine smoothing the field around it with a filter of scale R.
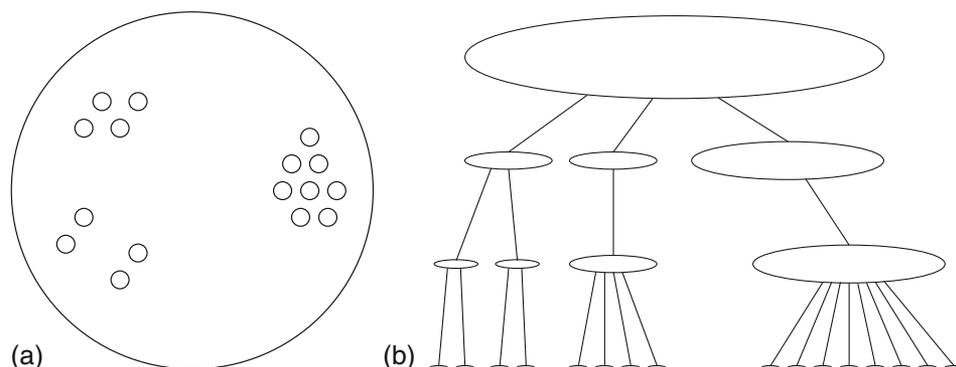
Figure 1. **(a)** Schematic drawing of the initial spatial distribution of objects that gives rise to the merger history tree shown in **(b)**. The largest circle represents the comoving size of the initial region associated with the final collapsed bound halo. As time evolves from the initial to the final, collapse time, this comoving radius decreases. The assumption is that all the matter initially within this region remains within it always. Thus, information about how the mass of a final object was partitioned into subhaloes at a given time contains information about the halo distribution smoothed on a scale given by the radius of the larger object at that time. **(b)** Schematic drawing of the associated merger history tree. Time increases upwards: the initial time is at the bottom of the figure. The branch on the right is associated with a region that, initially, was made up of many small objects that were close to each other, but rather separated from any other objects. The branch on the left, on the other hand, is associated with a region that was initially populated rather more homogenously.

As one changes R, the overdensity within the filter will change. Let a plot of the value of the smoothed density around this point be a function of R. For very large R (say, the Hubble volume), the overdensity in the smoothed filter should be negligible – the Universe is homogeneous on large scales. As R decreases, the value of the smoothed overdensity will vary, sometimes up, other times down. The jagged line in figure 2 shows that the result looks like a random walk – we shall shortly discuss whether the steps in the walk are truly independent. The x-axis is not quite R, but it is a monotonically decreasing function of R for reasons we discuss shortly. The y-axis shows the initial overdensity multiplied by the linear theory growth factor D_0/D_i.

The nonlinear physics of collapse is encoded in the initial overdensity which must be reached for an object to collapse and virialize by the present time. For a collapsing sphere, this critical overdensity δ_c is independent of the initial radius of the object [5]. This is peculiar to standard gravity: δ_c depends on this radius in many if not most modified gravity models [6]. In standard gravity, δ_c also depends on the initial shape of the object; this alters the details but not the logic of the excursion set approach [7].

Now, although any given walk starts from the origin, it will eventually reach the height δ_c. (This assumes there are fluctuations on arbitrarily small scales; which is true for ΛCDM models, but may not be true in general. Even in CDM models, objects smaller than the Jeans scale ($\sim 10^4 M_\odot$) may not form; we shall have more to say about this later.) This first crossing of δ_c (it may go on to cross δ_c many times at still smaller R) is significant: it
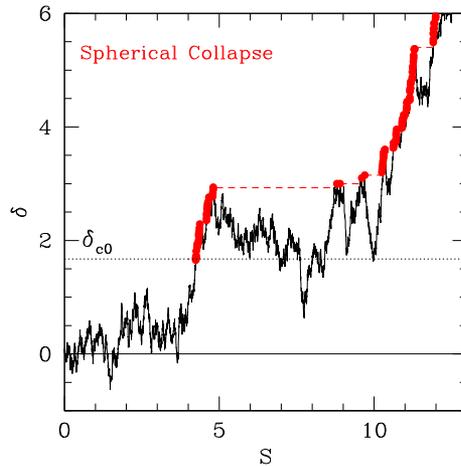
Figure 2. Mass history associated with a random position in the initial fluctuation field (jagged line), shown in scaled units: The critical density for spherical collapse (dotted line) decreases as time increases, and mass decreases as S increases. If one imagines sliding the dotted line downwards from great height, then filled circles show the pairs (S, δ) at which the walk would first cross this line. The horizonal jumps (connected by dashed lines) show places where S, and so the mass, changes dramatically – mergers (from [8]).

indicates that, when smoothed on this scale, the field was dense enough initially that it should have just collapsed and formed a virialized object today. In the spherical collapse model, shells do not cross, so the mass associated with this collapsed object is simply the mass that was originally within the smoothing filter R. Since all the fluctuations are small, this mass $M \propto R^3$. (This also shows why the subsequent crossings of δ_c are not so significant – their mass is included in M. It is only the first crossing which is significant.)

Moreover, in the spherical model, the critical density required for collapse at t is independent of mass, and this critical density is a decreasing function of time. The dotted line at $\delta \approx 1.686$ in figure 2 represents this critical value for t_0. At earlier times, this critical value was larger. The dots show the result of sliding a horizontal line downwards from great height, and recording the values of S at which the line first touches the walk. The set of (S, δ) values obtained in this way is actually a set of (M, t) values: this set can be thought of as describing the mass M of the collapsed object that this particle is in at time t. The figure shows that, in this model, the mass increases monotonically with time, but the increase in mass can sometimes be due to the rather large 'instantaneous' jumps.

In more picturesque language, this is a model of the mass history of objects, in which mass changes can be due to major or minor mergers, but the mass growth is hierarchical – there is no fragmentation. Moreover, for any given time, one can distinguish between the region to the right of the walk, which describes the formation 'history' of the object, and the region to the left, which describes its 'future'. Since the 'future' is actually the region surrounding the patch of interest, this approach makes specific statements about the joint distribution of the mass of an object, its formation history and its surrounding environment

[9,10]. In particular, this formulation provides a simple way of encoding the fact that a different large-scale environment evolves just like a Universe of an appropriately defined background density but of the same age (so the effective Hubble constant is also a function of environment) [11].

Now, clearly, the shape of the walk, and the scale R on which the walk first crosses δ_{sc}, and indeed, the whole set of (M, t) values, will change from one initial position or particle to another. If we imagine each object at time t as having been assembled by a sequence of mergers, then the whole set of walks associated with the various positions in the initial conditions contains information about the forest of all possible merger history trees. The excursion set ansatz is that statistical averages over this bundle of walks can provide information about various properties of this forest. The devil is in the details of exactly how one should calculate these statistical averages.

1.1 *The first crossing distribution in the excursion set approach*

The assumption that is most amenable to analytic progress is that the appropriate ensemble of walks is simply the one associated with randomly placed cells in the initial field. Then, the fraction of walks that first crosses δ_{sc} when the smoothing scale is R or greater equals the fraction of mass that is bound up in halos of mass greater than $M \propto R^3$:

$$\int_M^\infty dm \, \frac{dn(m|t)}{dm} \, \frac{m}{\rho} = \int_0^{\sigma(M)} d\sigma \, f[\sigma | \delta_c(t)]. \tag{1}$$

Similarly, consider the subset of walks which first crossed $\delta_{sc}(T)$ on scale R. For this subset, one can calculate the fraction of walks which first crosses $\delta_{sc}(t) > \delta_{sc}(T)$ on scales between r and R (note r must be smaller than R). The excursion set ansatz is that this equals the fraction of the total mass in clumps having M at time T that was in clumps of mass m or greater (of course, $m \leq M$) at the earlier time t. Hence, in this approach, the first crossing distribution is simply and directly related to the mass function of the collapsed objects.

For Gaussian initial conditions, different k modes are independent. Therefore, for a filter set which only allows one k-mode at a time, known as the sharp-k filter, different scales differ by the addition or subtraction of independent k-modes. For such a filter, the steps in the walk are truly independent, and the fraction of walks which crossed δ_c prior to σ is given by

$$\int_0^\sigma d\sigma \, f(\sigma | \delta_c) = 2 \int_{\delta_c}^\infty \frac{d\delta}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\delta^2}{2\sigma^2}\right) = \mathrm{erfc}\left(\frac{\delta_c}{\sqrt{2}\sigma}\right), \tag{2}$$

where the mirror-image symmetry-in-stochasticity argument, so nicely summarized in Chandrasekhar's review of 1943 [12], is responsible for the factor of 2 [1]. Differentiating with respect to σ gives the first crossing distribution:

$$f(\sigma^2 | \delta_c) \, d\sigma^2 = \frac{|\delta_c|}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\delta_c^2}{2\sigma^2}\right) \frac{d\sigma^2}{\sigma^2}. \tag{3}$$

This indicates that objects with mass greater than $m_*(t)$, where $\delta_c(t) = \sigma(m_*)$, are exponentially rare.

Notice that the left-hand side of eq. (2) → (1) as $\sigma \to \infty$: this in eq. (1) implies that, at any time (given by $\delta_c(t)$), all mass is in bound objects of some mass. (If one believes there is a lower limit to the mass which can collapse (e.g. the Jeans mass), or if there is no power at large k so that $\sigma(m)$ does not exceed a maximum value σ_{max} (e.g. Warm Dark Matter or neutralinos), then the total mass fraction in collapsed objects → erfc($\delta_c/\sqrt{2}\sigma_{max}$).)

Since different steps in the walk are independent, such a filter produces walks which are as stochastic as possible. Though it has received essentially no attention in the literature, the opposite limit, that of complete correlation, is also interesting. In this limit, a walk that has height $\nu = \delta/\sigma(R)$ on scale R has height $\delta/\sigma(r) = \nu$ on all other r also. In this case, the first crossing distribution is given by the expression above, but without the factor of 2. Hence, only half the mass is bound up in objects. This 'complete correlation' idea is a novel way of viewing the calculation of [13]; the factor of 2 difference comes from the fact that walks are not completely correlated, but exhibit some degree of stochasticity around the completely correlated case. As a result, stochasticity allows some of the walks which were otherwise prohibited from crossing δ_c to do so. This is a useful way to think of the effect of changing from the sharp-k filter to the real-space TopHat, for which steps are correlated over a range of scales. We shall return to this later.

Before moving on, it is worth noting that to apply the moral equivalent of the mirror-image argument to n-dimensional walks, the barrier must be an n-dimensional sphere. The solution may be related to questions about the paths taken by photons which reach us from the Sun, but I am not aware if Chandra worked on such generalizations of the formalism. Similarly, this method is particularly powerful for the small class of distributions that are said to be 'stable', in the sense that they keep their functional form under convolutions. The Gaussian is the best-known example, the Cauchy distribution is another example, and the Holtzmark distribution, also studied by Chandra, is the third. I would not be surprised to learn that Chandra worked out the first passage problem for these other distributions, even though he does not appear to have published the results. E.g., for the Cauchy distribution, $p(x)\mathrm{d}x = \mathrm{d}x/(1 + x^2)/\pi$ where $x = \delta/\sigma$, the first crossing distribution is given by replacing the erfc on the right-hand side of eq. (2) with $1 - (2/\pi)\arctan(\delta_c/\sigma)$. (N.B., whereas the 'rms' height of the walk grows in proportion to the square root of the number of steps for the Gaussian case, here it grows linearly.)

2. Formation times

It is conventional to define the 'formation time' of a halo as the first time that half the mass of the parent M-halo is assembled in subclumps of mass greater than m_{min}. In general, one might expect the shape of this distribution to depend on the value of m_{min}. In what follows, we shall show that the distribution is quite well fit by a lognormal for a range of values of m_{min}, and we provide a new excursion set calculation of the shape of this distribution. It is precisely this regime which many advocate as the one which determines the halo concentration today.

Our starting point is to notice that if $m_{min} \geq M/2$, then there can be only one subclump with the required mass; if such a subclump is present, then at least half the mass has certainly been assembled, and so the halo is said to have formed. In this case, the distribution of formation times can be estimated by the argument in [2]. (They actually studied the case

$m_{\min} = M/2$; Nusser and Sheth [14] provide an obvious generalization to $m_{\min} \geq M/2$.) The argument is that the probability that formation occurred at a higher redshift than z equals the probability that the mass in subclumps which are more massive than $m_{\min}$ at z exceeds one half. If $m_{\min} \geq M/2$, then there can be at most one subclump which satisfies this limit, and so the parent is said to have formed when a single subclump exceeds the mass limit. Therefore,

$$P_1(>z) = \int_z^\infty \mathrm{d}z' \, p(z') = \int_{m_{\min}}^M \mathrm{d}m_1 \, N(m_1|M, D_0), \tag{4}$$

where $N(m_1|M, D_0)$ is the mean number of m_1-subclumps identified at z_1 of an M-halo identified at z_0, and $D_0 \equiv \delta_1 - \delta_0$, where δ_1 is the critical density required for an initially overdense perturbation to collapse spherically and virialize at z_1, and similarly for δ_0. The actual formation time distribution $p(z)$ can be obtained by differentiating the right-hand side of this expression with respect to z. In the excursion set approach,

$$N(m|M, D) \, \mathrm{d}m = \frac{M}{m} \left(\frac{D^2}{s - S} \right)^{1/2} \frac{\mathrm{e}^{-D^2/2(s-S)}}{\sqrt{2\pi}(s - S)} \frac{\mathrm{d}s}{\mathrm{d}m} \, \mathrm{d}m, \tag{5}$$

where $s \equiv \sigma^2(m)$, σ denotes the r.m.s. value of the linear density fluctuation field when it is smoothed with a top hat filter of scale $R = (3m/4\pi\bar{\rho})^{1/3}$, and $S \equiv \sigma^2(M)$ is defined similarly [1,2]. This means that the expression for $p(z)$ above is actually a function of the variable $(\delta_1 - \delta_0)^2/(s_{\min} - S)$. Inserting eq. (5) in eq. (4), setting $m_{\min} = M/2$, and differentiating with respect to z yields

$$p_1(\omega) \, \mathrm{d}\omega = 2\omega \, \mathrm{erfc}\left(\frac{\omega}{\sqrt{2}} \right) \mathrm{d}\omega, \tag{6}$$

where $\omega^2 \equiv (\delta_{\mathrm{cf}} - \delta_{\mathrm{c0}})^2/(s_{\mathrm{f}} - S)$, $\delta_{\mathrm{cf}} = \delta_{\mathrm{c}}(z_{\mathrm{f}})$, and $s_{\mathrm{f}} = \sigma^2(M/2)$ [2].

Circles, hexagons, squares and triangles in figure 3 show this distribution of scaled formation times for haloes which have masses in the range 0.25–0.5, 1–2, 4–8, and 8–64M_* today. (The figure has been labelled with $\omega_{0.5}$ to emphasize the fact that the formation time is defined as the first time that half the mass has been assembled in pieces which are each at least half the mass at the present time.) The extent to which all the different symbols trace out the same curve is a measure of how weakly the scaled formation time depends on halo mass. The solid curves show eq. (6); it provides a reasonable description of the simulation results, although there is a weak tendency for halos to form at slightly earlier times than this formula predicts.

Given the reasonable success of eqs (4)–(6), it is useful to extend the argument beyond $m_{\min} \geq M/2$. To do so, suppose that $M/4 \leq m_{\min} \leq M/2$. Then there are two possibilities: formation occurred because half the mass was assembled in a single subclump of mass $\geq M/2$, or because there were two subclumps each with mass in the range $m_{\min} \leq m \leq M/2$. This means that

$$P_2(>z) = P_1(>z) + \frac{1}{2} \int_{m_{\min}}^{M/2} \mathrm{d}m_1 \, N(m_1|M, D_0)$$
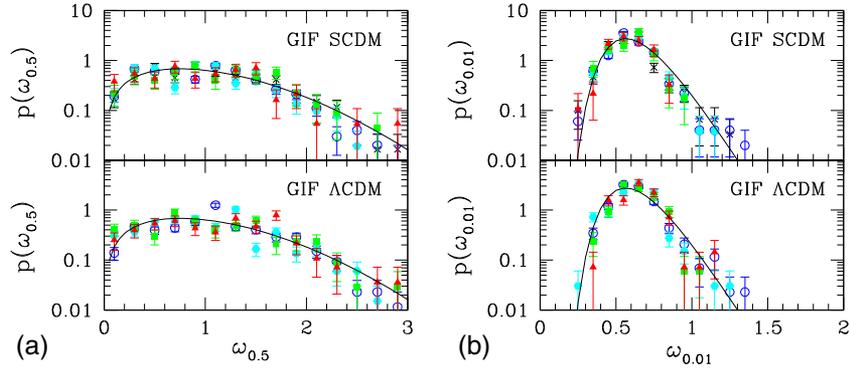$$\times \int_{m_{\min}}^{M/2} \mathrm{d}m_2 \, N(m_2|M - m_1, D_1), \tag{7}$$

Figure 3. Scaled distribution of formation times in the GIF simulations; formation is defined as the first time that half the total mass M is assembled in pieces more massive than $M/2$ **(a)** and $M/100$ **(b)**. In practice, this means that one subclump must contain at least half the final mass. Triangles, squares, hexagons and circles show the distribution for haloes which have masses M in the range 8–64, 4–8, 1–2, and 0.25–0.5M_* today. Solid line in **(a)** shows the analytic formula for this distribution. The scaled formation time distribution depends only weakly on halo mass. Solid line in **(b)** shows a lognormal distribution with mean $\ln \omega = \ln(0.6)$ and variance $\sigma_{\ln \omega} = 0.25$.

where $P_1(>z)$ is obtained from the final expression above, D_1 is defined by using $(\delta_1 - \delta_0)/(1-m_1/M)$ rather than $(\delta_1 - \delta_0)$ in the conditional mass function of m_2 given $(M-m_1)$ (see [15] for the reason why), and the factor $\frac{1}{2}$ comes from the fact that the subclumps are indistinguishable, and so the pair (m_1, m_2) should not be counted differently from the pair (m_2, m_1). The actual distribution of interest, $p(z)$, is obtained by differentiating $P_2(>z)$ with respect to z.

It is straightforward to extend this argument to the case $M/2^{n+1} \leq m_{\min} \leq M/2^n$ for any $n > 0$. Then

$$P_{n+1}(>z) = P_n(>z) + \int_{m_{\min}}^{M/2^n} dm_1 \cdots \int_{m_{\min}}^{M/2^n} dm_n$$

$$\times \frac{1}{n!} \prod_{i=1}^{n} N(m_i | R_{i-1}, D_{i-1}), \tag{8}$$

where $R_0 = M$, $R_j = M - \sum_{i=1}^{j} m_i$ for $j \geq 1$, and $D_i = (\delta_1 - \delta_0)/(R_i/M)$. This shows clearly that $P_n > P_{n-1}$ for all z, which reflects the fact that, in hierarchical models in which small clumps merge to make big clumps, formation can occur at higher redshifts as the limiting mass $m_{\min}$ is decreased. Note that the resulting distribution $p(z)$ will be a function of $(\delta_1 - \delta_0)^2/(s_{\min} - S)$.

When the initial power spectrum is white noise, then the integrand reduces to

$$\frac{(\nu^2/2\pi)^{n/2}}{n!} \int \prod_i^n \frac{d\mu_i}{\mu_i^{3/2}} \frac{\exp[-\nu^2(M/R_n - 1)/2]}{(R_n/M)^{3/2}}, \tag{9}$$

where $\nu^2 = (\delta_1 - \delta_0)^2/S_0$. This expression requires evaluation of an n-dimensional integral. The generic form of the integrals is an error function times a power law times a Gaussian. For example, if $m_{\mathrm{lim}} = M/3$, then the second term in $P_2(>z)$ can be written as

$$\int_{\nu_{\min}/2}^{\nu_{\min}} \frac{\mathrm{d}\nu}{\nu} \left(1 + \frac{\nu_{\min}}{\nu}\right) \sqrt{\frac{\nu}{2\pi}} \, \mathrm{e}^{-\nu/2} T\left(\nu + \nu_{\min}, \frac{\nu}{\nu + \nu_{\min}}\right),$$

where

$$T(\mu, \tau) = \sqrt{\frac{2\mu}{\pi}} \left(\frac{\mathrm{e}^{-\mu/(4-6\tau)}}{\sqrt{1/(2-3\tau)}} - \frac{\mathrm{e}^{-\mu/(2-4\tau)}}{\sqrt{1/(1-2\tau)}}\right)$$
$$- (1-\mu)\left[\mathrm{erf}\left(\sqrt{\frac{\mu/2}{2-3\tau}}\right) - \mathrm{erf}\left(\sqrt{\frac{\mu/2}{1-2\tau}}\right)\right],$$

and $\nu_{\min} \equiv (\delta_1 - \delta_0)^2/S$.

In practice, it is more efficient to solve the integrals numerically using a Monte-Carlo method. The idea is to generate an ensemble of merger history trees using the algorithm in [15], and then simply count the number of times that formation happened at z. The jagged curves in figure 4 show the result of doing this for $m_{\min} = M/n$, with $n = 2, 4, 8, 16$, and 32. The width of the distribution decreases as n increases. The smooth curves show our analytic formulae for the formation time distribution (eqs (4) and (7)), which correspond to $n = 2$ and 3, and we have shown $n = 4$ as well. The analytic curves pass through the Monte-Carlo ones demonstrating that our numerical Monte-Carlo approach works well. Notice that, for a wide range of $m_{\min}$, the resulting distribution is skewed, and so a lognormal should provide a reasonable approximation.
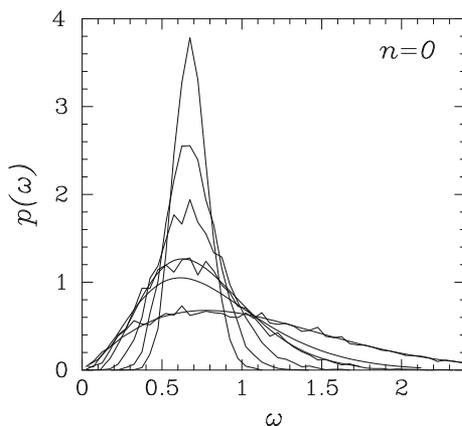

Figure 4. Dependence of scaled formation time distribution on definition of formation. An M-halo is said to have formed when half the mass is first assembled in pieces more massive than fM. Jagged curves show our Monte-Carlo calculation of the shape of this distribution when $f = 1/2$ (broadest distribution), 1/4, 1/8, 1/16 and 1/32 (narrowest distribution). Smooth curves show the distribution computed analytically for $f = 1/2$, 1/3 and 1/4.

We are now in a position to compare our formulae for the distribution of scaled formation times with the distribution measured in simulations. Figure 3b shows the formation time distribution when formation is defined as the first time that half the mass is assembled in pieces each of which are more massive than 0.01 times the final mass. (To emphasize this definition, the plot is labelled with $\omega_{0.01}$.) The solid line, which provides a reasonable description of the simulations, shows a lognormal distribution with mean $\ln\omega = \ln(0.6)$ and rms $\sigma_{\ln\omega} = 0.25$.

3. The Ornstein–Uhlenbeck process

The simplest successful models for predicting the abundances of dark matter halos make a specific assumption, known as the Markov assumption, about the correlation between density fluctuations on different scales. Amosov and Schücker [16] describe a non-Markov model for the halo distribution, and use it to interpret measurements of halo abundances. I show analytically and numerically that an error in the early parts of their analysis compromises all of their results. None of their conclusions are justified.

3.1 *Non-Markov random walks: Specific model*

In models of large-scale structure formation, the initial density fluctuation field is usually assumed to be Gaussian. This means that the amplitudes and phases of the Fourier waves which make up the field are independent of one another. The result of smoothing the field with a filter is another Gaussian field; a consequence of the fact that smoothing is a linear operation. Smoothing corresponds to a convolution in coordinate space, so it is a multiplication in Fourier space.

Let dB/ds denote standard Brownian motion, and let $B(t) = \int_0^t ds \, dB(s)/ds$. If we think of this as $B(t) = \int_0^\infty ds \, dB(s)/ds \, W(s, t)$, with $W = 1$ for $s \leq t$, and $W = 0$ otherwise, then it is convenient to think of dB/dt as representing the Fourier modes of a Gaussian field, and $B(t)$ as representing a smoothed version of the field, with smoothing scale proportional to $1/t$. This filter is special, in that the smoothed field $B(t)$ is independent of modes with $s > t$.

Another such smoothed field can be constructed from the Ornstein–Uhlenbeck process:

$$d\delta_T(t) = \frac{dt}{T} \int_0^t ds \, \frac{dB(s)}{ds} \, \exp\left(-\frac{t-s}{T}\right) \qquad (10)$$

for some $T \geq 0$. Pure Brownian motion corresponds to the limit in which $T \to 0$. For this model,

$$\delta_T(t) = \int_0^t ds \, \frac{dB(s)}{ds} \left[1 - \exp\left(-\frac{t-s}{T}\right)\right], \qquad (11)$$

which shows that the associated filter is $W = 1 - e^{-(t-s)/T}$ for $s \leq t$ and it is zero otherwise. In this model also, $\delta_T(t)$ depends only on the Brownian motion at $s \leq t$, but not on the Brownian motion beyond t. This is the model studied by [16].

Because it is built by summing Gaussians, the distribution of $\delta_T(t)$ is Gaussian with mean $\langle\delta_T(t)\rangle = 0$. The variance is obtained as follows. Since $B(t)$ is true random motion,
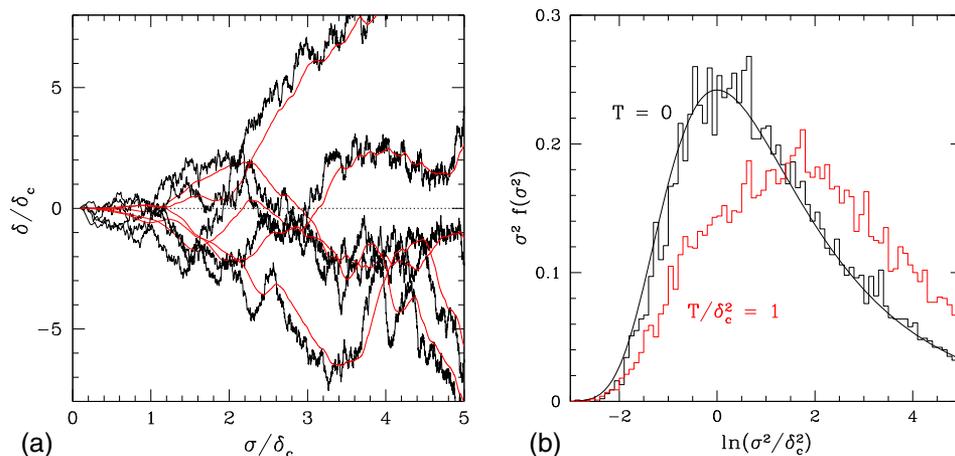
Figure 5. **(a)** Examples of non-Markov walks (smooth curves) constructed from Markov walks (jagged curves) following eq. (11). **(b)** Comparison of first crossing distributions associated with eq. (11) for $T = 0$ and δ_c^2. Smooth curve shows that eq. (3) provides a good fit only when $T = 0$.

the variance of $B(t)$ scales as $\langle B^2(t) \rangle \propto t$. For a general smoothing filter W, the variance is $\propto \int \mathrm{d}s\, W(s, t)^2$. So,

$$
\begin{aligned}
\langle \delta_T^2(t) \rangle &\propto \int_0^t \mathrm{d}s \left[1 - \exp\left(-\frac{t - s}{T} \right) \right]^2 \\
&\propto T \left(\frac{t}{T} - \frac{3}{2} + 2\mathrm{e}^{-t/T} - \frac{\mathrm{e}^{-2t/T}}{2} \right).
\end{aligned}
\tag{12}
$$

Physically motivated models of the smoothed density fluctuation field involve smoothing filters that typically introduce correlations between all steps of the Brownian motion. For instance, a Gaussian smoothing filter has $\delta_G(t) = \int_0^\infty \mathrm{d}B(s)\, \exp(-s^2/2t^2)$. Since this filter falls rapidly at $s \gg t$, filters which ignore the contribution from $s \geq t$ may not be a bad approximation. This is the main motivation for studying the Ornstein–Uhlenbeck-based process of eqs (10) and (11) further.

The expressions above explicitly show how to construct a random walk $\delta_T(t)$ from the underlying Brownian motion. Examples of such walks, and the underlying Brownian walks from which they were constructed, are shown in figure 5 (smooth and jagged trajectories, respectively): T was set equal to δ_c^2, and results are shown scaled by δ_c.

3.2 *First crossing distribution: Numerical*

In the excursion set model [3], the quantity which is related to the halo mass function is the distribution of first crossings of a barrier of height $\delta_c(\sigma)$ by the random walks $\delta(\sigma)$. In the simplest such model, δ_c is a constant independent of σ. In the units of figure 5, this is the line $\delta/\delta_c = 1$.

Amosov and Schücker [16] assert that the first crossing distribution $f(t, \delta_c)$ is a function of $\delta_c/\sigma_T(t)$ only, where $\sigma_T^2(t) = \langle \delta_T^2(t) \rangle$ of eq. (12): all dependence on T comes from the dependence of σ on T. In particular, because $T = 0$ is pure Brownian motion, they assert that the first crossing distribution has the same form as eq. (3).

To test this prediction, I generated an ensemble of Brownian motion random walks. From these, I constructed the ensemble of walks associated with different values of T. I then used these to generate numerically the first crossing distributions of B and δ. Figure 5b shows the simulated first crossing distributions as functions of σ_T/δ_c for $T = 0$ and $T = \delta_c^2$, with $\sigma_T(t)$ given by eq. (12). The smooth curve shows eq. (3). It provides a good fit to the $T = 0$ case, as it should, because this is pure Brownian motion. The agreement between simulation and theory suggests that my numerical algorithm is accurate. However, the simulated first crossing distribution for $T = \delta_c^2$ is very different from that for $T = 0$.

3.3 *Mirror image method: Invalidity*

The numerical experiments show that, when expressed as a function of δ_c/σ_T, the form of the first crossing distribution depends strongly on T, contrary to the assertion of [16]. They used what they call the 'standard' mirror image method to arrive at their solution. But, as I show below, this argument is incorrect for walks such as those described by eq. (11). The argument is as follows:

Let

$$F(t, \delta_c) = \int_t^\infty \mathrm{d}s \, f(s, \delta_c), \tag{13}$$

and let $p(\delta_T, t)$ denote the probability that the walk has the value δ_T at t. Note that $p(\delta_T, t)$ is Gaussian with mean $\langle \delta(t) \rangle = 0$ and variance given by eq. (12). Then

$$F(t, \delta_c) = P(<\delta_c, t) - \int_0^t \mathrm{d}s \, f(s, \delta_c) \, P(<\delta_c, t|\delta_c, s), \tag{14}$$

where $P(<\delta_c, t) = \int_{-\infty}^{\delta_c} \mathrm{d}\delta_T \, p(\delta_T, t)$ and similarly for $P(<\delta_c, t|\delta_c, s)$. The first term counts all the walks that are below δ_c at $t > s$, and the second subtracts off those walks which had crossed δ_c at some $s \leq t$, and then walked to some value less than δ_c at t. The trick is to evaluate the second term correctly. Amosov and Schücker [16] set this term equal to $P(>\delta_c, t)$ and invoke the mirror image argument to justify their choice.

The mirror image argument assumes that for every walk that continues upwards from δ_c at $s < t$, there is an equally likely trajectory which follows from reversing the signs of all the steps between s and t. This cannot be true if $p(\delta_T, t|\delta_c, s)$ does not describe a symmetric distribution with mean $\langle \delta_T(t)|\delta_c, s \rangle = \delta_c$. For walks described by eq. (11),

$$\delta_T(t) = \left(1 - \mathrm{e}^{-(t-s)/T}\right) B(s) + \mathrm{e}^{-(t-s)/T} \delta(s) + \delta(t - s), \tag{15}$$

where $t > s$ and $\delta_T(t - s)$ are given by an expression like (11), but with the understanding that the steps $\mathrm{d}B$ are those from s to t. If it is known that $\delta_T(s) = \delta_c$, then $p(\delta_T, t|\delta_c, s)$ is Gaussian with mean

$$\begin{aligned}
\langle \delta_T(t)|\delta_c, s \rangle &= \left(1 - \mathrm{e}^{-(t-s)/T}\right) \langle B(s)|\delta_T(s) = \delta_c \rangle \\
&\quad + \mathrm{e}^{-(t-s)/T} \delta_c
\end{aligned} \tag{16}$$

and variance

$$\langle(\delta_T(t) - \langle\delta_T(t)|\delta_c, s\rangle)^2 |\delta_c, s\rangle = \sigma_T^2(t) - \sigma_T^2(s)\frac{\langle\delta_T(t)|\delta_c, s\rangle^2}{\delta_c^2},\tag{17}$$

where we have used the fact that the mean of eq. (11) is zero. Since

$$\begin{aligned}\langle B(s)|\delta_T(s) = \delta_c\rangle &= \delta_T(s)\frac{\int_0^s dx\,[1 - e^{-(s-x)/T}]}{\langle\delta_T^2(s)\rangle}\\ &= \delta_c\frac{s/T - [1 - e^{-s/T}]}{s/T - 3/2 + 2e^{-s/T} - e^{-2s/T}/2}\end{aligned}\tag{18}$$

the term in angle brackets on the right-hand side of eq. (16) does not, in general, equal δ_c. Therefore, the left-hand side is, in general, a function of t. Therefore, the mean is different from δ_c, and the steps from s to t are not distributed symmetrically about δ_c. Hence, the mirror image argument does not apply. Blind use of this argument will lead to an incorrect expression for the first crossing distribution.

Inserting eq. (18) for $\langle B|\delta\rangle$ into the right-hand side of eq. (16) shows that $\langle\delta_T(t)|\delta_c, s\rangle$ equals δ_c times a term that depends on s/T and $(t - s)/T$. A little algebra shows that for all choices of $T > 0$ and $t > s$, this term is larger than unity. Therefore, the mean $\delta_T(t)$ given $\delta_T(s) = \delta_c$ is greater than what Amosov and Schücker [16] assumed. So their expression overestimates the true value of the second term in eq. (14). This means that they are incorrectly subtracting too much from the first term, and so their expression for $F(t, \delta_c)$ is too small at all t for all $T > 0$. This explains the sense of the discrepancy in the numerical experiments (see figure 5b).

This analysis has a curious implication. Recall that $\langle\delta_T(t)|\delta_T(s) = \delta_c\rangle$ is proportional to δ_c. Therefore, if $\delta_c = 0$, then the walk beyond s is symmetric about $\delta_c = 0$, and the mirror image argument can be applied. Hence, eq. (3) is the distribution of times for walks which begin at $-\delta_c$ and first reach 0 at t.

4. Broken symmetry

4.1 *Dependence on smoothing filter: Correlated steps*

Whether or not the field, when smoothed on scale R, lies above δ_c, depends on the smoothing filter. To illustrate this, we have chosen two filters: one is a tophat sphere in configuration-space, and is the filter most closely associated with the physics of collapse; the other is a tophat in k-space (so it oscillates in configuration-space). Figure 6a shows how the mean density around a random position of height δ_c depends on the smoothing filter, for different choices of the underlying power spectrum $P(k) \propto k^n$ (the expressions we plot can be derived in a straightforward manner from results in [17]). Except for the case of white noise ($n = 0$), when both filters are equivalent, the tophat filter (dashed line) has a larger mean value and smaller scatter around the mean, compared to the sharp-k filter (solid line). As a result, relationship between σ and M depends on smoothing filter. And, more importantly, the fraction of walks which first crosses $\delta_c(t)$ on some scale $\sigma(M)$ also depends on how one smooths the underlying field.

The following discussion provides some insight into the form of the first crossing distribution associated with these filters. Recall that, for the sharp-k filter, the first crossing
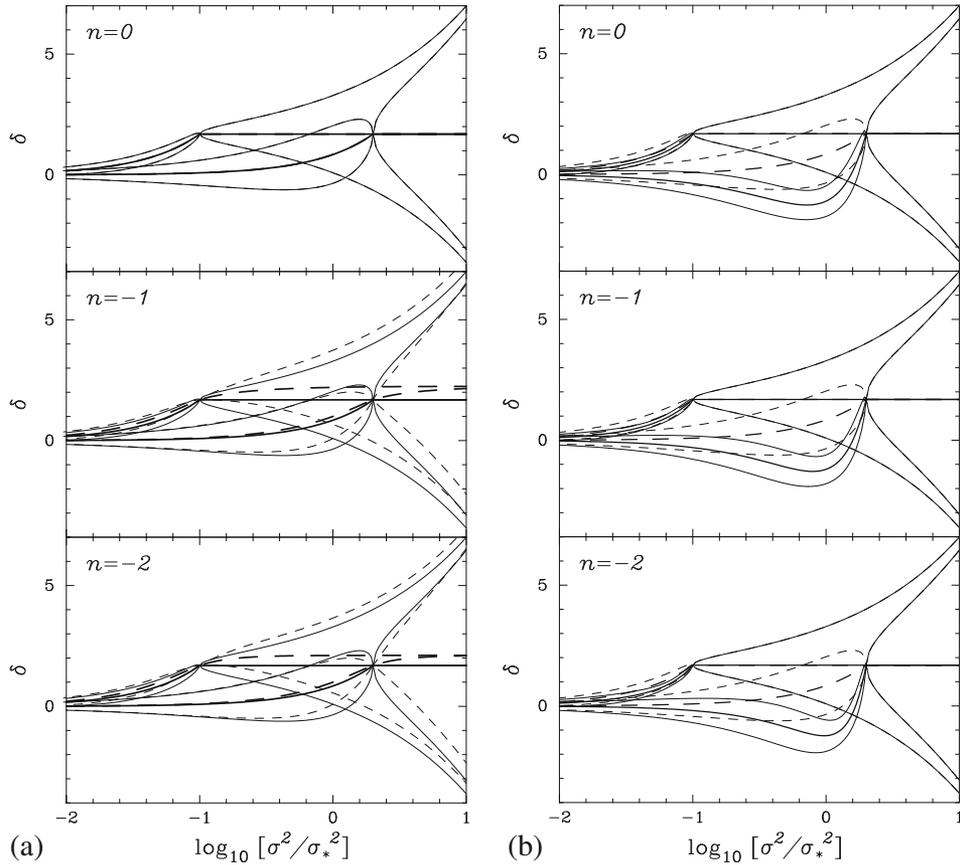
Figure 6. **(a)** Comparison of the mean and rms density run within and around a random position, when smoothed with a Gaussian filter (dashed curves) and one which is sharp in k-space (solid curves). Results for two positions are shown: the curves which intersect at small σ/σ_* are for regions defined to have height δ_c identified on a larger smoothing scale (we have set $\sigma_* \equiv \delta_c$). **(b)** Comparison of these quantities for a peak in the k-space field (solid curves), and a randomly chosen position that happens to have the same height on that scale (dashed curves). On average, on scales larger than the peak scale, the peak density profile lies below that for a random position of the same height, and the scatter around this mean profile is narrower, although the difference between peaks and random positions is less pronounced for the higher ($\delta_c \gg \sigma$) peaks which might be associated with massive halos.

distribution is given by eq. (2), but that this produces walks which are, in some sense, as stochastic as possible. The first crossing distribution of completely smooth correlated walks differs from that for k-space filtering by a factor of 2. The TopHat filter is less stochastic than the k-space filter, but it is not completely smooth/deterministic either. As a result, it will be rather like the completely correlated case (i.e., no factor of two in eq. (2)) for small values of σ, but the fact that it is not completely correlated means it will depart from this
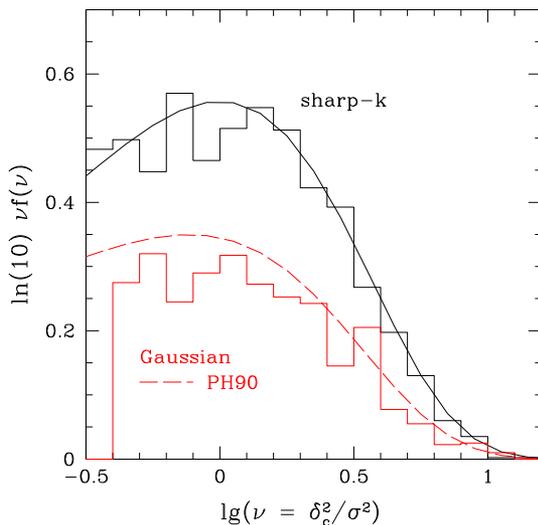
Figure 7. First crossing distribution of a barrier of constant height δ_c by walks with independent steps (upper histogram), and with correlated steps, in which the correlations are due to a Gaussian smoothing filter on a field with $P(k) \propto k^{-1}$ (lower histogram). Curves show exact (upper) and approximate (lower) expressions for the associated first crossing distributions.

at intermediate scales. Because of its simplicity, it is the *k*-space-related formulae which have guided the work over the last 15 years or so, despite the fact that reasonably good approximations for the case of correlated steps are available [18]. Figure 7 illustrates this.

4.2 *Peaks and the excursion set approach: Which ensemble of walks?*

A significant drawback of the approach above is the assumption that the appropriate ensemble of walks over which to average is that associated with random positions in the field. For example, it is not unreasonable to assume that the most massive objects today were associated with peaks in the initial field. Since positions which are peaks (on a certain scale) are a special subset of all positions, the statistics of peaks are clearly different from those of randomly placed cells [17]. Figure 6b shows why these differences are likely to be important. The dashed curves are the same as the solid curves of figure 6a (sorry!) and the solid curves show the corresponding profiles for (*k*-space smoothed) peaks of the same height. Clearly, the correlation between the formation history of an object of a given mass, and its surrounding environment, will depend crucially on whether or not it is formed from a peak. The importance of this effect is just beginning to be explored.

4.3 *One more step in the excursion set approach: Correlated walks*

As an alternative to worrying about the statistics of peaks, progress can also be made if one modifies eq. (1) for the relation between the first crossing distribution and the mass

function. For instance, consider a walk whose first crossing distribution predicts mass m. At the very least, one would like to ensure that all the other walks that are within R_m of this one predict smaller masses; and that this walk itself is further than R_M from all walks for which the predicted mass is $M > m$. Incorporating this effect is a tough but interesting open problem, for which a crude estimate can be given as follows.

Let $\phi(m)$ denotes the quantity which should be on the right-hand side of eq. (1). If $p(M|m)$ denotes the probability that a walk which was predicted to have mass m actually ends up in a halo of mass M, then

$$\phi(M) = f(M) + \int_0^M dm\, f(m)p(M|m) - f(M) \int_M^\infty dM'\, p(M'|M), \quad (19)$$

the second term counts the increase in the abundance of M because of this effect, and the third counts the decrease, as, for similar reasons, objects originally predicted to have mass M are assigned to more massive objects. Rearranging the order of the integrals in the second term shows that

$$\phi(>M) = f(>M) + \int_0^M dm\, f(m) \int_M^\infty dM'\, p(M'|m) \quad (20)$$

(note that $M' > m$ surely). Since all quantities in the final term on the right-hand side are positive, ϕ will be shifted towards higher mass scales than f.

Another way of arriving at the same result is to assume that the initial conditions can be divided up into a set of disjoint regions, each of which contains the mass which will make up one halo at a later time (say, the present). Each such patch contains some mass: let $\phi(>M)$ denotes the mass fraction associated with regions containing mass greater than M (i.e., it equals the left-hand side of eq. (1)). Associated with each patch is a bundle of random walks (the number of walks being proportional to M), and associated with each walk is a predicted mass m which comes from applying the excursion set algorithm to the walk. If we assume that the 'centre' of each region is a local maximum of the predicted mass [7], then $p(M'|m)$ in eq. (20) is the probability that the mass of a region is M' when the excursion set predicts m (note that $M' > m$ surely).

This means that the first crossing distribution is related to $\phi(M)$ by

$$f(m) = \int_m^\infty dM\, \phi(M)\phi(m|M), \quad (21)$$

where $\phi(m|M)$ denotes the probability that a randomly chosen walk which lies within an M-patch has predicted mass m. One may think of this as the fraction of the volume of M which predicts m. Further insight can be obtained by noting that if $f(m)$ is given by the first crossing distribution with barrier height δ_c, then the expression above is satisfied if $\phi(M)$ is given by the first crossing distribution with barrier height $a\delta_c$ for some $a < 1$, and $\phi(m|M)$ is the fraction of walks which starts at $(a\delta_c, M)$ and first crosses δ_c on scale m. Thus, a crude model for the difference between the first crossing distribution f, and the mass fraction in halos ϕ, is that ϕ looks like f but with a smaller value of δ_c. In other words, although the physics of collapse yields a barrier of height δ_c, the result of doing the appropriate statistical averaging makes it look like the barrier is lower. So it is remarkable that just such a rescaling, with $a \sim 0.85$, appears to be necessary for the excursion set approach to provide a good description of halo abundances measured in simulations [7].

Ravi K Sheth

The work in progress shows that this formalism also provides a simple way to model the effects of different smoothing filters. So it is the natural language with which to describe the effects of both correlated steps (the effect of the smoothing filter) and correlated walks (the effect of spatial correlations which modify the appropriate ensemble of walks over which to average).

Acknowledgements

References

 [1] J R Bond, S Cole, G Efstathiou and N Kaiser, *Astrophys. J.* **379**, 440 (1991)
 [2] C Lacey and S Cole, *Mon. Not. R. Astron. Soc.* **262**, 627 (1993)
 [3] J Shen, T Abel, H J Mo and R K Sheth, *Astrophys. J.* **645**, 783 (2006)
 [4] R K Sheth, *Mon. Not. R. Astron. Soc.* **300**, 1057 (1998)
 [5] J E Gunn and J R Gott, *Astrophys. J.* **176**, 1 (1972)
 [6] M C Martino, F Stabenau and R K Sheth, *Phys. Rev.* **D79**, 084013 (2009)
 [7] R K Sheth, H J Mo and G Tormen, *Mon. Not. R. Astron. Soc.* **323**, 1 (2001)
 [8] J Moreno, C Giocoli and R K Sheth, *Mon. Not. R. Astron. Soc.* **397**, 299 (2009)
 [9] H J Mo and S D M White, *Mon. Not. R. Astron. Soc.* **282**, 347 (1996)
[10] R K Sheth and G Tormen, *Mon. Not. R. Astron. Soc.* **329**, 61 (2002)
[11] M C Martino and R K Sheth, *Mon. Not. R. Astron. Soc.* **394**, 2109 (2009)
[12] S Chandrasekhar, *Rev. Mod. Phys.* **15**, 1 (1943)
[13] W Press and P Schechter, *Astrophys. J.* **187**, 425 (1974)
[14] A Nusser and R K Sheth, *Mon. Not. R. Astron. Soc.* **303**, 685 (1999)
[15] R K Sheth and G Lemson, *Mon. Not. R. Astron. Soc.* **305**, 946 (1999)
[16] G Amosov and P Schücker, *Astron. Astrophys.* **421**, 425 (2004)
[17] J M Bardeen, J R Bond, N Kaiser and A S Szalay, *Astrophys. J.* **304**, 15 (1986)
[18] J A Peacock and A F Heavens, *Mon. Not. R. Astron. Soc.* **243**, 133 (1990)