Halo abundances and counts-in-cells: the excursion set approach with correlated steps

Aseem Paranjape,1⋆ Tsz Yan Lam2⋆ and Ravi K. Sheth1,3

1The Abdus Salam International Center for Theoretical Physics, Strada Costiera 11, 34151 Trieste, Italy
2IPMU, University of Tokyo, Kashiwa, Chiba 277-8583, Japan
3Center for Particle Cosmology, University of Pennsylvania, 209 South 33rd Street, Philadelphia, PA 19104, USA

Accepted 2011 November 2. Received 2011 October 26; in original form 2011 May 11

ABSTRACT
The excursion set approach has been used to make predictions for a number of interesting quantities in studies of non-linear hierarchical clustering. These include the halo mass function, halo merger rates, halo formation times and masses, halo clustering, analogous quantities for voids and the distribution of dark matter counts in randomly placed cells. The approach assumes that all these quantities can be mapped to problems involving the first-crossing distribution of a suitably chosen barrier by random walks. Most analytic expressions for these distributions ignore the fact that, although different $k$-modes in the initial Gaussian field are uncorrelated, this is not true in real space: the values of the density field at a given spatial position, when smoothed on different real-space scales, are correlated in a non-trivial way. As a result, the problem is to estimate first crossing distribution by random walks having correlated rather than uncorrelated steps. In 1990, Peacock & Heavens presented a simple approximation for the first crossing distribution of a single barrier of constant height by walks with correlated steps. We show that their approximation can be thought of as a correction to the distribution associated with what we call smooth completely correlated walks. We then use this insight to extend their approach to treat moving barriers, as well as walks that are constrained to pass through a certain point before crossing the barrier. For the latter, we show that a simple rescaling, inspired by bivariate Gaussian statistics, of the unconditional first crossing distribution, accurately describes the conditional distribution, independent of the choice of analytical prescription for the former. In all cases, comparison with Monte Carlo solutions of the problem shows reasonably good agreement. This represents the first explicit demonstration of the accuracy of an analytic treatment of all these aspects of the correlated steps problem. While our main focus is on first crossing distributions of deterministic barriers by random walks, in Appendices we also discuss several issues that arise upon introducing a stochasticity in the barrier height, a topic which has gained interest recently with regard to the mapping between first crossing distributions and halo mass functions.

Key words: large-scale structure of Universe.

1 INTRODUCTION
The abundance of clusters and its evolution is a useful probe of the primordial fluctuation field, the subsequent expansion history of the universe and the nature of gravity. This is, in part, because there is an analytic framework for understanding how cluster formation and evolution depends on the background cosmological model. Analyses based on the assumption that clusters form from a spherical collapse suggest that a cluster today is a region that is about 200 times the background density, and it formed from the collapse of a sufficiently overdense region in the initial conditions (Gunn & Gott 1972). Numerical simulations suggest that this expectation is reasonably accurate.

Hence, the problem of estimating cluster abundances at any given time reduces to the problem of estimating the abundance of sufficiently overdense regions in the initial conditions (Press & Schechter 1974). However, the overdensity associated with a given position in space depends on scale (in homogeneous cosmologies, the likely range of overdensities is smaller on large scales). So, to estimate cluster abundances, the problem is to find those regions in the initial conditions which are sufficiently overdense on a given
smoothing scale, but not on a larger scale. This is because, if the larger region is sufficiently overdense, then, as it pulls itself together against the expansion of the background universe and collapses, it will also squeeze the regions within it to smaller and smaller sizes. The framework for not double counting the smaller overdense regions that are embedded in larger ones is known as the excursion set approach (Epstein 1983; Bond et al. 1991; Lacey & Cole 1993). With some care, one can build an excursion set model for voids as well (Sheth & van de Weygaert 2004).

Within the context of the excursion set approach, the spherical evolution model exhibits an important technical simplification: the critical overdensity $\delta_c$ required for collapse at a given time is independent of the mass or size of the final object. In fact, neither clusters nor voids are spherical, and the critical overdensity associated with collapsed objects depends on how different from spherical the object is. As a result, the critical overdensity required for collapse, when averaged over objects of a given mass, becomes mass dependent (Sheth, Mo & Tormen 2001). In modified gravity models, mass dependence of the critical overdensity appears even in spherical evolution models (Martino, Stabenau & Sheth 2009; Brax, Rosenfeld & Steer 2010).

While mass dependence of $\delta_c$ does not complicate the logical framework of the excursion set description, it does impact the ability to obtain exact analytic expressions for the quantities of interest. Nevertheless, simple accurate approximations have been developed (Sheth & Tormen 2002; Lam & Sheth 2009). In addition to allowing one to predict cluster abundances, these allow one to produce accurate excursion set models which describe how the probability distribution function of mass in randomly placed cells depends on cell size (Sheth 1998; Lam & Sheth 2008). In some respects, clusters can be thought of as cells of vanishingly small size, so the excursion set model for the counts-in-cells distribution is also a model for the density run around clusters (and voids) on scales that are larger than the virial radius (or void wall).

To obtain analytic expressions, all of these analyses assume that the density field on one scale is trivially correlated with that on another scale. If one plots the overdensity as a function of smoothing scale, then this resembles a random walk—the usual assumption is that successive steps in the walk are independent of the previous ones. This is known to be a bad approximation, but because there are no known exact solutions to the case of realistic correlations, the assumption of uncorrelated steps has been routine. This is despite the fact that Peacock & Heavens (1990) showed how to derive a reasonably accurate expression for the spherical collapse problem and correlated steps. However, it has received little use, presumably because it is only an approximation, whereas the corresponding problem of spherical collapse with uncorrelated steps was solved exactly shortly after their paper appeared (Bond et al. 1991). Bond et al. also described a simple numerical solution to the correlated steps problem; it too has received little attention.

Recently, however, there has been renewed interest in the correlated steps problem: Maggiore & Riotto (2010a) have introduced field theory techniques to address this problem. In essence, this approach aims to solve analytically the same path integrals that Bond et al. (1991) solved numerically. This technical machinery is too complicated to solve exactly, but, at the end of the day, it does provide a simple analytic approximation for cluster abundances in the spherical collapse model. There is as yet no similarly simple expression for the case in which $\delta_c$ is mass dependent (although see De Simone, Maggiore & Riotto 2011a for a treatment of correlations induced by non-Gaussian initial conditions).

This paper is motivated by the fact that the field theoretic approach yields approximate rather than exact expressions. So, it is interesting to ask how it compares to the older Peacock–Heavens approximation. In Section 2, we show that the older approximation is, in fact, the more accurate of the two, for the case of a constant, deterministic barrier (see below for a discussion of the case when the barrier height is stochastic). Therefore, the main goal of the current paper is to show how the analysis of Peacock & Heavens can be extended to the case of mass-dependent $\delta_c$. Section 3 shows that this can be done almost trivially. Section 4 discusses an ansatz which relates the shape of the first crossing distribution of walks that are constrained to pass through a certain point before first crossing the barrier to the shape of the unconditional first crossing distribution. We combine this with the Peacock–Heavens approach to provide a rather accurate approximation of the conditional distribution. A final section summarizes and discusses some implications. Appendix A provides an alternative derivation of the Peacock–Heavens ansatz which yields some insight into the nature of their approximation, while Appendix B contains technical details about smoothing windows. Appendix C includes a discussion of problems in which the barrier height is stochastic, an issue which has gained considerable recent interest (Maggiore & Riotto 2010b; Corasaniti & Achitouv 2011b). We argue that these latter treatments correspond to making some specific technical choices which are difficult to test, and that there are in fact several other (testable) options when dealing with stochastic barriers, which remain to be explored.

In a related paper, we address the question of halo bias (Paranjape & Sheth 2011), which is associated with a certain limit of our solution of the constrained walks problem. And, in a third paper, we address the question of voids (Paranjape, Lam & Sheth 2011): here, the problem is the generalization of the Peacock–Heavens approximation to the case of two absorbing barriers rather than just one. This also works rather well. All of these analyses assume that the initial fluctuation field was Gaussian. The case of non-Gaussian initial conditions is discussed by Musso & Paranjape (2011).

2 THE CONSTANT BARRIER PROBLEM

In what follows, it will be useful to define

$$\sigma^2_j(R) \equiv \int \frac{dk}{k} \frac{k^3 P(k)}{2\pi^2} k^{3j} W^2(kR),$$

(1)

where $P(k)$ is the power spectrum of initial density fluctuations (linearly extrapolated to present epoch) and $W$ is a smoothing filter. The quantity $\sigma^2_j(R)$ measures the variance in the field on scale $R$. We will reserve the symbol $s$ (or $S$) to denote the variance

$$s \equiv \sigma^2_0(R).$$

(2)

We will also make use of the combination

$$\gamma \equiv \frac{\sigma^2_1}{\sigma^2_0}.$$  

(3)

For $P(k) \propto k^n$ and a Gaussian filter $W(kR) = e^{-(kR)^2/2}$, which we will use to illustrate many of our results, $\sigma^2_j \propto R^{3n-2j-2}$ and $\gamma^2 = (3 + n)/(5 + n)$.

The excursion set ansatz relates the abundance of haloes of mass $m$ to the fraction of random walks which first cross a barrier of height $\delta_c$ on scale $s(R)$, where $m = \bar{\rho} 4\pi R^3/3$. If $f(s) ds$ denotes this fraction, then

$$\frac{m}{\bar{\rho}} \frac{dm}{dn} = f(s) ds.$$  

(4)
We argue elsewhere that this relation between halo abundances and the first crossing distribution is not the full story. In what follows, we are mainly interested in making accurate estimates of the first crossing distribution.

2.1 Numerical (Monte Carlo) solution: tophat smoothing and LCDM $P(k)$

Fig. 1 shows the result of the Monte Carlo simulation of the first crossing distribution associated with a barrier of constant height $\delta_c$. In the top panel, the black histogram (the one which has more counts at large $y = \delta_c^2/s$) shows the Monte Carlo distribution for walks with uncorrelated steps. The solid curve going through it shows the associated analytic expression for the first crossing distribution (equation 6). The agreement indicates that the numerical algorithm works.

The red histogram shows the result when the steps are correlated. In practice, we transformed each walk with uncorrelated steps into one with correlations by applying smoothing filters of different scales following Bond et al. (1991). In this case, the correlation depends on the form of the filter and on the shape of the initial linear theory power spectrum $P(k)$. We used a tophat smoothing filter and a cold dark matter ($\Lambda$CDM) power spectrum appropriate for $(\Omega_m = 0.25, \Omega_\Lambda = 1 - \Omega_m, h = 0.7$ and $\sigma_8 = 0.8$). We then performed the same analysis as for the uncorrelated walks: find and store the scale on which $\delta_c$ is first crossed. Note that now the relation between $s$ and $M$ is modified compared to the previous case; we have checked that our algorithm does this correctly.

The dotted curve shows the first crossing distribution associated with what we call completed correlated walks below (equation 8), a limiting case that will prove useful for understanding many of the results to follow. The short and long dashed curves show the Peacock–Heavens approximation (equation 17), and the dot–dashed curve shows the approximation from equation (119) of Maggiore & Riotto (2010a). Both approximations have one free parameter, which we have set to the value appropriate for the walks shown in Fig. 1. For Maggiore–Riotto, this parameter is $\kappa = 0.45$. For Peacock–Heavens, the parameter $\Gamma$ is actually scale dependent (see Fig. 4 below); the short dashed curve shows the result of setting $\Gamma = 4$ and ignoring this dependence; the long-dashed curve, which provides a better description of the numerical solution, includes this dependence. (For reasons that we discuss later, we calculated $\Gamma$ for the long-dashed curve assuming a Gaussian rather than tophat filter.)

The bottom panel of Fig. 1 shows the residuals (red squares) with Poisson errors between the tophat CDM walks and the Peacock–Heavens approximation with scale-dependent $\Gamma$, and the relative difference between the Maggiore–Riotto result and the Peacock–Heavens approximation (dashed blue curve). For comparison, we also show the residuals between the walks with uncorrelated steps and the corresponding analytic prediction from Bond et al. (1991; black triangles). The points for uncorrelated steps were given a small horizontal offset for clarity.

This shows that the Peacock–Heavens approximation for the first crossing distribution is more accurate than that of Maggiore & Riotto.1 In what follows, we will show that it is also more easily extended to treat more general power spectra, more general smoothing filters and more general barrier-crossing problems. (In private communications with us, Maggiore & Riotto have emphasized that they do not believe that equation 119 of Maggiore & Riotto 2010a can be applied directly, without further computation, to the case of more general smoothing filters or power spectra.)

2.2 Power-law $P(k)$ and other smoothing filters

Fig. 2 shows the first crossing distribution of a constant barrier when $P(k) \propto k^{-1.2}$. From top to bottom, the histograms show results for sharp-$k$ (black), tophat (blue) and Gaussian (red) smoothing filters.

1 The latter shows discrepancies at the level of $\sim 15$ per cent, in keeping with the fact that their treatment is a linearization in $\kappa \sim 0.4$. 

**Figure 1.** Top panel: distribution of the scale $s$ on which walks which first cross $\delta_c$ (histograms) for a LCDM $P(k)$. The histogram which shows more objects at large $y = \delta_c^2/s$ is from walks with uncorrelated steps (labelled sharp-$k$); the other shows the result for walks with correlated steps due to tophat smoothing. The solid curve shows the analytic prediction for uncorrelated steps from Bond et al. (1991); the dotted curve is this divided by a factor of 2, for completely correlated steps; the two dashed curves show two implementations of the Peacock & Heavens (1990) approximation for a Gaussian filter (see the text for details); and the dot–dashed curve shows the approximation of Maggiore & Riotto (2010a). Bottom panel: red squares show the residuals (with Poisson errors) between the walks with correlated steps and the scale-dependent implementation of the Peacock–Heavens approximation. The blue dashed curve shows the relative difference between the Maggiore–Riotto and Peacock–Heavens results. For comparison, the black triangles (which were given a small horizontal offset for clarity) show the residuals between the walks with uncorrelated steps and the corresponding analytic prediction of Bond et al. (1991).
Walks with uncorrelated steps are jagged and stochastic. For such walks, the first crossing distribution is

\[ f_u(s) ds = -\frac{\partial P_u(s)}{\partial s} = -\frac{\partial}{\partial s} \text{erf}(\delta_c/\sqrt{2s}), \]

(5)

\[ = \frac{\delta_c}{\sqrt{2\pi s}} e^{-\delta_c^2/2s} \, ds \]

(6)

(Chandrasekhar 1943; Bond et al. 1991), where \( P_u(s) \) is the ‘survival probability’ that a randomly chosen walk has not crossed the barrier \( \delta_c \) prior to \( s \).

Walks with correlated steps are smoother. However, before stating the Peacock–Heavens approximation for walks with correlated steps, we believe it is useful to study another case which can be solved exactly and which is in some sense the opposite of the uncorrelated steps problem. In this subsection, we will be interested in the limit in which the walks are as smooth and deterministic as possible. In particular, we would like to think of such walks as having completely correlated steps, where by complete correlation we mean that the height of the walk at one time completely specifies its value at all other times.

This notion of a completely correlated walk is, at first sight, somewhat ambiguous, since a walk which has the same height \( \delta \) at all \( s \) could be said to be completely correlated, but a walk which is defined by a straight line from the origin through the point in question \((\delta, s)\), or indeed, any curve \( f(s) \) whose value is completely specified by the pair \((\delta, s)\), would also be completely correlated. However, if we now require that the ensemble of such completely correlated walks also satisfy the constraint that the fraction of walks which lie above \( \delta \) on scale \( s \) equals \( \text{erf}(\delta/\sqrt{2s})/2 \), i.e. obeys Gaussian statistics, then it must be that the number which specifies each member of this ensemble is \( v = \delta/\sqrt{s} \). Thus, a \( v = 1 \) walk is one which has height \( \delta = \sqrt{s} \) on scale \( s \); a \( v = 2 \) walk is one which has height \( 2\sqrt{s} \) on scale \( s \), etc. Note that these are walks which are not constant height \( \delta \), but constant \( v \): their height scales as \( \sqrt{s} \). We will refer to this family of walks as having completely correlated steps.

Now consider a barrier of constant height \( \delta_c \). The first crossing distribution associated with this family of walks will depend on the distribution of \( v \). If this distribution is Gaussian, \( p(v) = \exp(-v^2/2)/\sqrt{2\pi} \), then the corresponding survival probability is

\[ P_v(s) = \frac{1}{2} \left( 1 + \text{erf}(\delta_c/\sqrt{2s}) \right). \]

(7)

As a result,

\[ s_f(s) = -\frac{\partial P_v(s)}{\partial \ln s} = -\frac{1}{2} \frac{\partial}{\partial \ln s} \frac{P_v(s)}{2} = s_f(s); \]

(8)

does this differs from the case of completely uncorrelated steps by the factor of 2 in the denominator. The origin of this factor is clear: if \( \delta_c > 0 \), then walks having \( v < 0 \) can never cross \( \delta_c \). For a Gaussian distribution of \( v \), this means half the walks never cross \( \delta_c \). This is a novel way to understand just what it is that Press & Schechter (1974) derived: their expression describes the first crossing distribution of
walks having completely correlated steps (in the sense described above).

Before moving on, note that it is trivial to extend this analysis to ‘moving’ barriers, whose height depends monotonically on $s$. For barriers $B(s)$ which decrease with $s$, one simply replaces $\delta_c \rightarrow B(s)$ in the expression above,

$$P_c(s) = \frac{1}{2} \left( 1 + \text{erf}(B(s)/\sqrt{2s}) \right).$$

For barriers which increase with $s$, this is slightly more involved, as shown by Fig. 3, which plots $\delta_c/\sqrt{s}$ versus $s$.

Complete correlated walks correspond to horizontal lines on this plot. It is then obvious that for barriers $B(s)$ such that $B(s)/\sqrt{s}$ has a minimum at some $s = S_{\text{crit}}$, there is no crossing of the barrier for $s > S_{\text{crit}}$. [Since $B(s)$ is monotonically increasing, $B(s)/\sqrt{s}$ monotonically increases for $s > S_{\text{crit}}$.] As a result, the survival probability becomes

$$P_c(s) = \begin{cases} \frac{1}{2} \left( 1 + \text{erf}(B(s)/\sqrt{2s}) \right), & \text{if } s \leq S_{\text{crit}}, \\ \frac{1}{2} \left( 1 + \text{erf}(B(S_{\text{crit}})/\sqrt{2S_{\text{crit}}}) \right), & \text{if } s > S_{\text{crit}}, \end{cases}$$

and the rate $f(s)$ is given by equation (10) until $s = S_{\text{crit}}$, at which point it becomes and stays zero. We will return to this point later.

To end this subsection, we note that an interesting example of an increasing barrier is one which increases as the square root of $s$:

$$B(s) = \delta_c (1 + \beta \sqrt{s}/\delta_c).$$

In this case, although completely correlated walks of height $\nu < \beta$ will never cross $B(s)$, we have $S_{\text{crit}} \rightarrow \infty$, so the complication associated with $s > S_{\text{crit}}$ does not arise. This barrier is a convenient approximation to the ‘moving barrier’ associated with haloes which form from an ellipsoidal collapse (Moreno, Giocoli & Sheth 2009).

While showing non-trivial $s$ dependence, it retains the simplicity of the constant barrier (see Fig. 3). The rate of increase of this barrier ($\sim s^{1/2}$) also serves as the dividing line beyond which $S_{\text{crit}}$ takes finite values and the complexity of equation (11) comes into play. Shen et al. (2006) argue that this form, with $\beta < 0$, is also a convenient approximation to the barrier associated with the formation of sheets.

### 2.4 Correlated, but not deterministic, steps

The two limiting cases – of maximally stochastic and completely deterministic walks – serve as useful guides for the construction of the first crossing distribution when there is some, but not complete, correlation between steps. For example, for the case of very weak or very strong correlations, one might imagine perturbing around one or another of these solutions. We will argue below that the Peacock–Heavens approximation may be thought of as perturbing around the case of complete correlation.

We first briefly restate their approximation, following its presentation in Bond et al. (1991). The Peacock–Heavens approximation derives from noting that the height of a walk on scale $s$ is correlated with its height on scales that are within its $\delta_{\text{f}}$, where $\Gamma$ is a parameter that depends on details of the filter, and will be defined below. One then asserts that the walk can be broken up into independent segments of length $s_{\text{f}}$ and requires that the height of the walk is below the barrier after each step. It may be helpful to think of this approximation as stating that, of the fraction of walks that are below $\delta_{\text{c}}$ after $n$ steps, one must take the fraction that were also below after $n - 1$ steps, and the fraction of these that were below after $n - 2$ steps and so on. If we define

$$c(\delta s) \equiv \int_{\delta s}^{\delta_{\text{f}} / \sqrt{s}} \frac{\exp(-x^2/2)}{\sqrt{2\pi}} \, dx,$$

then the Peacock–Heavens ansatz for the survival probability in the presence of a single barrier of height $\delta_{\text{c}}$ is

$$P_{\text{PH}}(s_n) = c(\delta_{\text{c}} | s_n) \prod_{i=1}^{n-1} c(\delta_{\text{c}} | s_i),$$

where the spacing between $s_i$ is chosen such that each step is approximately independent of the previous ones. One then expresses the product as the exponential of a sum over logs, and then replaces the sum by an integral. If we define

$$p(s) = c(\delta_{\text{c}} | s),$$

and

$$E_c(s) \equiv \exp \left( \int_0^{s_{\text{f}}} \frac{dv}{\Gamma^2 v^2} \ln p(s) \right),$$

then

$$s_{\text{f}} p(s) \equiv \frac{\partial P_{\text{PH}}(s)}{\partial \ln s} = \frac{\partial \left[ p(s) E_c(s) \right]}{\partial \ln s} = E_c(s) \left( s_{f} p(s) - \frac{p(s) \ln p(s)}{\Gamma} \right),$$

where $f$ was defined in equation (8). Appendix A discusses why this is only an approximation to the exact solution.

What remains is to determine $\Gamma$. But before doing so, note that as $\Gamma \rightarrow \infty$, then the term in the exponential of equation (16) → 0, so $E_c$ itself → 1, and in equation (17) $f(p(s)) \rightarrow f_c(s)$, the distribution for completely correlated walks. Hence, if we view $E_c(S)$ as a series in $1/\Gamma$, then one may think of the Peacock–Heavens approximation as perturbing around the case of smooth completely correlated walks. In fact, we can rewrite the survival probability as

$$P_{\text{PH}}(s) = P_c(s) E_c(s).$$
S were defined in equation (1). This can be written = \Gamma_1 n / (we have assumed → (19) 2 / \Gamma_1 > − E 1. The tophat is well behaved with the CDM s = ϵ B β s / δ C s 1 B k R c ln(2) > S = σ 0.5 P 1429–1441 2 s is a delta function, centred on = x = 0 on scale (420, 0, the resulting expression for the first crossing distribu- s 1.5 2 P < P (has a not /Γ_1 γ − P). ι is well defined in this

\[
Γ = 2 \pi \ln(2) \sqrt{γ^2/(1 − γ^2)}. \tag{19}
\]
To see what this implies, suppose that P(k) ∝ k^n. Then, γ^2 = (3 + n)/(5 + n), making Γ = 2 \pi \ln(2) \sqrt{(3 + n)/2}. For n = (−1.2, −2), Γ = (4.13, 3.08). For a more realistic CDM spectrum, Fig. 4 shows that Γ varies with scale (approximately logarithmically with s), with similar numerical values. But if the filter is a tophat in real space, then some of the integrals appearing in equation (19) may diverge, so one must compute Γ slightly differently. In Appendix B, we show that for power-law spectra Γ = 2 \pi \ln(2) \sqrt{(n + 1)/(n + 3)}/(n − 3), so the Peacock–Heavens procedure is well defined only if −3 < n < −1. Note that the tophat filter is known to be analytically troublesome with power-law spectra, with even the variance s being undefined for n ≥ 1. The tophat is well behaved with the CDM spectrum though, and Fig. 4 shows that Γ is well defined in this case. Nevertheless, as Figs 1 and 2 show, the numerical results for tophat filtering are actually better described by the analytical approximation for Gaussian filtering.

The oscillatory behaviour of the Fourier transform of the tophat filter results in tophat smoothed walks being less correlated than Gaussian-smoothed walks. One therefore expects that the value of Γ for the tophat filter should be smaller than that for the Gaussian. Numerically, since the first crossing distributions for the two filters are not very different, the difference in values of Γ should at most be a factor of the order of unity, and this is true for the CDM spectrum. The dramatic discrepancy for power-law spectra is most likely caused by the known bad behaviour of the tophat in this case, combined with the fact that the Peacock–Heavens procedure is a perturbation around the completely correlated case, and therefore performs better for Gaussian-smoothed walks, which are closer to complete correlation than are tophat walks.

One possibility for alleviating the problem of the tophat filter with power-law spectra is to round the sharp edge of the tophat (of scale R_TH) with, say, a Gaussian filter of a smaller scale ε R_TH (Bond et al. 1991). The resulting Γ is well defined for all n > −3, and smaller than the Gaussian. Fig. 4 shows the result for a CDM spectrum, using ε = 0.1. In the remainder of the paper, we will use the Peacock–Heavens analysis for Gaussian smoothing and power-law spectra, since this combination is simple, well defined and accurate. The discussion above (and a comparison of Figs 1 and 2) shows that, with some care, our conclusions can be generalized to more realistic power spectra and filters.

3 MOVING BARRIERS
One of the virtues of the Peacock–Heavens approximation is that it is easy to see how equation (18) should be extended to the case when δ is no longer constant. The logic behind writing the correction factor E_c(s) as in equation (16) does not depend on the barrier being constant. Therefore, for any moving barrier B(s), this suggests setting 

\[
p(s) = c (\epsilon < B(s)) / \sqrt{5}, \tag{20}
\]
in equation (16).

On the other hand, the discussion in Section 2.3 shows that the survival probability for completely correlated walks P_c(s) must be handled with care, depending on the nature of the moving barrier. For barriers B(s) which are decreasing functions of s, we can still make the simple replacement δ_c → B(s) in equation (7). Fig. 5 shows that, for barriers of the form

\[
B(s) = \delta_c (1 + β s / δ_c^2) \tag{21}
\]
with β < 0, the resulting expression for the first crossing distribution (dashed curve) describes the Monte Carlo solution rather well. We also note that, for this case, the expression for completely correlated walks in equation (10) (dotted curve) actually describes the numerical solution more accurately. We return to this in Section 5.

For barriers which increase with s, such that B(s)/S > 3 has a minimum at some s = S_min, Section 2.3 shows that we must use equation (11) rather than equation (7) to calculate the completely correlated survival probability P_c(s) used in equation (18). This correctly accounts for walks that never cross the barrier. Fig. 5 shows that this prescription works well in describing the Monte Carlo solution for a linearly increasing barrier (21) with β > 0, although there is a discrepancy at small y. Note that in this case, the completely correlated expression does not perform well.

4 CONstrained WALKS
The Peacock–Heavens approximation can also be extended to describe the first crossing distribution of walks which are conditioned to pass through some non-zero (δ, S). In many respects, this problem highlights the difference between walks with uncorrelated steps and those which are completely correlated.

Since a completely correlated walk is specified by a single number, if it is known to have height δ_c on scale S, then it will have height δ_S = δ_c / S on scale s. That is, the expected distribution of heights on scale s is a delta function, centred on δ_S / S. On the other hand, for uncorrelated walks, this distribution is Gaussian with mean δ_S and variance s − S (we have assumed s > S). This means that, except for a shift of origin, the conditional walk follows the same statistics
as the unconditioned one. Namely, one simply sets $\delta_c \rightarrow \delta_c - \delta_S$ and $s \rightarrow s - S$ in equation (6) for the first crossing distribution.

Note that the deterministic walk increases its height by an amount

$$\delta_c (\sqrt{s} - S)$$

whereas stochastic walks have no change on average, but some walks will reach large heights because the rms increases as $\sqrt{s} - S$. The general case will lie somewhere in between these two extremes: we might generically expect a milder increase in the expected height compared to the deterministic case, with a narrower distribution around the mean compared to the stochastic case.

### 4.1 A scaling ansatz based on bivariate statistics

To extend the Peacock–Heavens approach to describe constrained walks, we will make an ansatz which we justify later. Our ansatz is that the conditional distribution is just a rescaled version of the unconditional one, where the scaling variable is inspired by bivariate Gaussian statistics. Namely, we define

$$v_{10} = \frac{\delta_c - r \delta_S \sqrt{s} S}{\sqrt{s(1 - r^2)}} = \frac{\delta_c - (S_c / S) \delta_S}{\sqrt{s - (S_c / S)^2 S}},$$

(22)

where $r = S_c / \sqrt{s} S$ and

$$S_c = \int \frac{dk}{k} \frac{k^2 P(k)}{2\pi^2} W(k R_c) W(k R_S).$$

(23)

Note that the integral above is similar to that which defines $s$ and $S$, the only difference being that here the two smoothing filters have different scales. Our ansatz is that, when expressed as $v_{10}(v_{10})$, the conditional first crossing distribution will have the same shape as the unconditional distribution $v_f(v)$, a point we will return to shortly.

Equation (22) accounts for the fact that if the field is constrained to have value $\delta_s$ on scale $S$, then its value on scale $s$ will be distributed around a mean value of $(S_c / S) \delta_S$. For uncorrelated steps $S_c = S$, so $s(1 - r^2) = s - S$ and hence $v_{10} = (\delta_c - \delta_S) / \sqrt{s - S}$. This corresponds to simply shifting the origin of the walk from $(0, 0)$ to $(\delta_S, S)$, as expected from the previous discussion. Completely correlated walks have $r = 1$; in this limit, our expression for $v_{10}$ correctly indicates no scatter around a mean value of $\delta_S \sqrt{s} / S$.

For Gaussian filtering of a power-law spectrum $r = [2 R_c R_S / (R_c^2 + R_S^2)]^{(n+3)/2}$ and $S \propto R_S^{(n+3)}$, making $r \sqrt{s} S = [2(1 + (S_c / S)^2(3+a))]^{(n+3)/2}$, and $s(1 - r^2) = s - S(2(1 + (S_c / S)^2(3+a))^{(n+3)})$. To get some intuition into what this implies, it is helpful to consider the limit $R_c \ll R_S$. In this case, $S \ll s$, so $v_{10} \approx (\delta_c - 2(3/2 \delta_S)) / \sqrt{s - 2^{3+n} S}$. This corresponds to a shift of origin that is larger than for the uncorrelated case, but much smaller than for the completely correlated case, and a variance that is slightly smaller than for the uncorrelated case, but much larger than for the completely correlated case. The discussion above applies for other smoothing filters too, except that the numerical coefficient $2^{(n+3)/2}$ will change. In Appendix B we give some details for the tophat filter.

### 4.2 Comparison with Monte Carlos

For the Peacock–Heavens approximation, our ansatz means that we set

$$c(< \delta_c | s) \rightarrow c(< \delta_c, s | \delta_S) = \frac{1}{2} \left[ 1 + \text{erf}(v_{10} / \sqrt{2}) \right]$$

(24)

in equation (15) when computing equations (16)–(18). Fig. 6 shows that this works reasonably well. The histograms show the first crossing distribution of a barrier of height $\delta_c$ by the subset of walks which are conditioned to pass through $\delta_S$, for a few choices of $\delta_c$, and the smooth curves show our extension of the Peacock–Heavens approximation. Although the qualitative trend is easy to understand – walks which start closer to the barrier (i.e. large $\delta_c$) cross it after fewer steps so there are few left to first cross at $s \gg S$ – our approach does a reasonable job of quantifying the effect.

Our ansatz is that the conditional distribution is just a rescaled version of the unconditional one. We have tested it more directly...
κ is small but not zero, the same as 0.5.

In an earlier version, their expressions (B29)–(B31) contained an error, as can be seen by requiring consistency with Ma et al. (2011) or by simply demanding that the limit \( S_0 \rightarrow 0, \delta_0 \rightarrow 0 \) reproduce the Maggiore–Riotto unconditional distribution. The factor \( \delta_0 - \delta_b \) should be \( \delta_b \) in the last line of each of equations (B29)–(B31). The error was corrected in a subsequent version of the paper, after it was pointed out by us.

Our ansatz, which allows one to transform any unconditional distribution into a conditional one (the accuracy of the resulting curve will be limited by that of the unconditional distribution, of course), allows us to provide some insight into the recent work of De Simone, Maggiore & Riotto (2011b), who present a path-integral analysis of the constrained walk problem. Their equation (B30)\(^2\) with \( S_0 \rightarrow 0 \) (derived using the path integral machinery after linearizing in \( \kappa \)) is, in fact, equivalent to simply replacing \( \nu \rightarrow \delta_0^2(1 + \kappa)/\sqrt{S} \) in their expression for the unconditional mass function, and then only keeping terms up to linear order in \( \kappa \). Since their \( \kappa \) is the \( \mathcal{S} \rightarrow 0 \) limit of our \( \mathcal{S}/\mathcal{S} - 1 \), we conclude that our simple ansatz of \( \nu \rightarrow \nu_{10} \) reproduces the detailed path integral result in the limit where path-integral calculations have been performed (i.e. to linear order in \( \kappa \)). Moreover, when \( \mathcal{S} \) is small but not zero, Figs 6 and 7 show that the same ansatz correctly describes the numerical solution, without having to linearize in \( \mathcal{S}/\mathcal{S} - 1 \) (which, in any case, is not small compared to unity). Of course, in our ansatz, \( \mathcal{S}/\mathcal{S} \) depends on scale and appears in the denominator of \( \nu_{10} \) as well.

### 4.3 Extensions and generalizations

Our treatment of the first crossing of \( \delta_0 \) by walks which are constrained to pass through some \( (\delta_0, \mathcal{S}) \) is easily extended in two ways.

\(^2\)In an earlier version, their expressions (B29)–(B31) contained an error, as can be seen by requiring consistency with Ma et al. (2011) or by simply demanding that the limit \( S_0 \rightarrow 0, \delta_0 \rightarrow 0 \) reproduce the Maggiore–Riotto unconditional distribution. The factor \( \delta_0 - \delta_b \) should be \( \delta_b \) in the last line of each of equations (B29)–(B31). The error was corrected in a subsequent version of the paper, after it was pointed out by us.

---

**Figure 6.** First crossing distribution of a barrier of height \( \delta_0 \) by the subset of walks which are conditioned to pass through \( (\delta_0, \mathcal{S}_0) \), for a few choices of \( \delta_0 \) (as labelled). Short dashed, solid and long-dashed curves show the analytic prediction from our extension of the Peacock & Heavens (1990) approximation (equation 24 in equation 17), for Gaussian smoothing of a Gaussian field with \( P(k) \propto k^{-1.2} \).

**Figure 7.** Conditional first crossing distributions of Fig 6, expressed as a function of the scaling variable \( y_{10} \equiv \nu_{10}^2 \) (equation 22). Symbols with error bars show the unconditional distribution from Fig. 2.
One is to the ‘two-barrier’ problem: given that the walk first crossed the barrier $\delta_{0}$ on scale $S_{0}$, what is the probability of first crossing the barrier $\delta_{1}$ on scale $S_{1}$? In this case, one uses $c(<\delta_{1}|s) \rightarrow c(<\delta_{0}|s_{0})$ for $S_{0} \ll s < S_{1}$, but $c(<\delta_{1}|s) \rightarrow c(<\delta_{0}|s_{0})$ for $s < S_{0}$. It will be interesting to test how well this simple extension does, compared to a numerical solution. In principle, we could go on to estimate merger rates from the limit in which $\delta_{0} \rightarrow \delta_{1}$ (following Lacey & Cole 1993). In this limit, however, most walks will cross the barrier within a few steps. As a result, most crossings will happen before the walk has travelled a distance that is of the order of the correlation length $\Gamma \delta$, so the Peacock–Heavens approximation is no longer expected to work.

The second generalization is the introduction of more constraints on the walks before (or after) they cross $\delta_{c}$. Again, to incorporate these, one must simply modify equations (24) and (22) appropriately, since the net effect of these additional constraints will be to change the mean and variance of the typical walk height on scales that are different from the constrained scales.

5 DISCUSSION

We have presented an analysis of the first crossing distribution of a moving barrier by random walks having correlated steps – the first analysis to explicitly compare analytic approximations with numerical (Monte Carlo) solutions. For walks with uncorrelated steps, exact solutions are only known for a handful of special cases, although a good analytic approximation is available for the general case (Lam & Sheth 2009). However, in the limit in which steps are completely correlated, so the walks are smooth and deterministic rather than jagged and stochastic, we showed that this solution is straightforward, at least for barriers which are monotonic functions of time/scale (Section 2.3).

For the more general case of some, but not complete, correlation between steps, Peacock & Heavens (1990) provide a simple approximation for the first crossing distribution of a barrier of constant height (equation 17). We showed that their approximation can be thought of as a correction to the solution for completely correlated steps (Section 2.4). The correlation involves a suitably defined correlation length $\Gamma$ (equation 19), whose inverse can actually be thought of as an expansion parameter; complete correlation is the limit $1/\Gamma \rightarrow 0$. While the prescription itself does not involve a perturbation in $1/\Gamma$, one can understand that the approach works well (Fig. 2) because $1/\Gamma \approx 0.2$ (Fig. 4). We argued that this is also why the approach works better for Gaussian than for tophat smoothed walks – the latter have smaller values of $\Gamma$.

We also showed (Fig. 1) that, when using a constant barrier height without scatter, the older Peacock–Heavens approximation was more accurate than that of Maggiore & Riotto (2010a). This more recent work, based on field-theoretic methods, explicitly perturbs around the solution for completely uncorrelated steps. In this case, the perturbation parameter $\kappa$, which should be $\ll 1$, is actually of the order of $1/2$. Since $\kappa$ is larger than one would like, and $\Gamma$ is smaller, it is certainly interesting to attack the problem of walks with correlated steps from both directions. (See e.g. Corasaniti & Achitouv 2011a for a recent extension of the path-integral approach to moving barriers.) Additionally, as we discuss below and in Appendix C, it is also interesting to explore the accuracy of these approaches when extended to models in which the barrier height shows a scale-dependent scatter rather than being deterministic.

In Section 3, we showed how to extend the Peacock–Heavens approximation to handle moving barriers. For barriers which decrease with time/scale, this extension is remarkably simple and remarkably accurate (Fig. 5). For barriers which increase sufficiently steeply with time/scale, the extension is slightly more complex, but still quite accurate. We summarize our results here: for a barrier $B(s)$ which is monotonic in $s$, the first crossing distribution is the derivative of the survival probability $f(s) = -\partial_{s}P(s)$, where

$$P(s) = P_{c}(s)E_{c}(s).$$

Here, $P_{c}(s)$ is the survival probability for completely correlated walks, given by equation (11), in which $S_{crit}$ is finite only when the barrier increases faster than $\approx \sqrt{s}$, and in particular $S_{crit} \rightarrow \infty$ for constant and decreasing barriers. The correction factor $E_{c}(s)$ is given by

$$E_{c}(s) = \exp \left( \int_{0}^{s} \frac{ds'}{\Gamma s'} \ln c(<B(s')|s') \right).$$

where $c(<\delta|s)$ was defined in equation (13).

Although this extension to moving barriers is relatively straightforward – it is far simpler than for walks with uncorrelated steps, or for the field theoretic approach – our comparison with Monte Carlo simulations indicates that there is room for improvement. In particular, Fig. 5 showed that, for barriers which decrease sufficiently rapidly with $s$ (for linear barriers, this is true for negative enough values of $\beta$), the solution associated with completely correlated walks is an excellent approximation. That is, the correction factor $E_{c}(s)$ should simply be set to unity. It is easy to see why this happens: for barriers which fall steeply – where steep means $B(s)$ changes height by more than $\sqrt{s}$ over the correlation length-scale $\Gamma$ – the barrier is almost vertical, and so one need not worry about walks which cross the barrier more than once. In this sense, the walks behave as though they are completely correlated. We are currently investigating whether our simple extension of the Peacock–Heavens ansatz can be improved to make $E_{c}$ depend, not just on the correlation length-scale $\Gamma$, but on how $\Gamma$ relates to the scale barrier shape.

In Section 4, we described a simple ansatz which allows one to transform the first crossing distribution of unconditional walks into a rather good estimate of the first crossing distribution associated with walks which are constrained to pass through a certain point in the $(\delta, S)$-plane before crossing the barrier. When applied to the Peacock–Heavens approximation, our ansatz for the conditional distribution (equations 24 and 22) works rather well (Fig. 6). As models for the unconditional distribution improve, we expect our ansatz to continue to provide a useful approximation, because the numerical solution does appear to scale as predicted (Fig. 7).

The simplicity of our approach to the problem of conditioned walks derives from approximating the problem, which potentially involves $n$-point distributions, to bivariate distributions. To impose more than one constraint one must simply modify equations (24) and (22); this is a straightforward generalization that we did not explore further, but is clearly a straightforward way to incorporate assembly bias effects of the sort identified by Sheth & Tormen (2004), and since studied by a number of authors.

We were careful to state at the beginning that our primary interest is in how well one can describe the first crossing distribution when steps are correlated. This is because, although our results provide increased understanding of halo abundances and evolution, a number of issues must be addressed before they can be used to provide quantitative constraints on cosmological parameters.

First, the correlated walk problem is known to underpredict the abundances of clusters (Bond et al. 1991). There are at least three possible resolutions: (i) the spherical model for collapse dynamics is wrong – $\delta_{c}$ must be smaller, or factors other than the initial...
overdensity matter (following e.g. Sheth et al. 2001, for clusters); (ii) there is substantial scatter around the actual value of $\delta_c$ – in which case one must decide whether the appropriate solution is to make the barriers associated with the void-in-void and void-in-cloud problems fuzzy or stochastic (following Sheth et al. 2001 or Maggiore & Riotti 2010b, respectively, and also see discussion in appendix B of Sheth & Tormen 2002), or whether it is better to simply convolve our solution for fixed $\delta_c$ with a distribution $p(\delta_c)$ of values (e.g. Appendix C); (iii) in addition to accounting for correlated steps, we must account for the fact that the appropriate ensemble of walks over which to average also contains correlations – in effect, our calculation has assumed each walk is a potential halo centre, whereas only a few walks really are (compare figs 2 and 3 in Sheth et al. 2001). For points (ii) and (iii) at least, this means that the fundamental assumption of the approach, equation (4), is incorrect.

In addition, the solution to the two-barrier problem associated for walks with uncorrelated steps is known to give a better description of halo merger histories than is the one for correlated steps (Bond et al. 1991). Thus, even though we have shown how to extend the Peacock–Heavens approach so as to provide a reasonable description of the random walk problem, we will now have to account for at least one of the three issues discussed above, point (iii) in particular, before we can claim to have a better model for halo formation. In the present case, this problem is twofold, since it is easy to see that the Peacock–Heavens approximation will break when the barrier to be crossed is low, since then the assumption that the walk can travel a distance $s\Gamma$ without crossing the barrier is no longer accurate. Further study along these lines is in progress.

ACKNOWLEDGMENTS

TYL was supported by World Premier International Research Center Initiative (WPI Initiative), MEXT, Japan, and JSPS travel grant (Institutional Program for Young Researcher Overseas Visits), and is grateful to the Center for Particle Cosmology at the University of Pennsylvania for hospitality. RKS is supported in part by NSF-AST 0908241.

REFERENCES

Bond J. R., Cole S., Efstathiou G., Kaiser N., 1991, ApJ, 379, 440
Brax P., Rosenfeld R., steer D. A., 2010, J. Cosmol. Astropart. Phys., 8, 33
Breiman L., 1966, Proc. 5th Berk. Symp. Math. Statist. Prob., 2, 9
Chandrasekhar S., 1943, Rev. Mod. Phys., 15, 1
Corasaniti P. S., Achitouv I., 2011a, Phys. Rev. Lett., 106, 241302
Corasaniti P. S., Achitouv I., 2011b, Phys. Rev. D, 84, 023009
De Simone A., Maggiore M., Riotto A., 2011a, MNRAS, 412, 2587
De Simone A., Maggiore M., Riotto A., 2011b, MNRAS, in press (arXiv:1102.0046)
Epstein R. I., 1983, MNRAS, 205, 267
Gunn J. E., Gott J. R., III, 1972, ApJ, 176, 1
Lacey C., Cole S., 1993, MNRAS, 262, 627
Lam T. Y., Sheth R. K., 2008, MNRAS, 386, 407
Lam T. Y., Sheth R. K., 2009, MNRAS, 398, 2143
Ma C.-P., Maggiore M., Riotto A., Zhang J., 2011, MNRAS, 411, 2644
Maggiore M., Riotto A., 2010a, ApJ, 711, 907
Maggiore M., Riotto A., 2010b, ApJ, 717, 515
Martino M., Stabnau H. F., Sheth R. K., 2009, Phys. Rev. D, 79, 084013
Moreno J., Giocoli C., Sheth R. K., 2009, MNRAS, 397, 299
Musso M., Paranjape A., 2011, MNRAS, submitted (arXiv:1108.0565)
Paranjape A., Sheth R. K., 2011, MNRAS, in press (arXiv:1105.2261)

APPENDIX A: APPROXIMATION

For a random walk with $n$ discrete time steps $\{s_i\}$, the joint probability distribution for the sequence of heights $\{\delta_i\}$ is a multivariate Gaussian with covariance matrix $\mathbf{C}_{ij} = \langle \delta_i \delta_j \rangle$. The probability that, after each of the $n$ steps, the height of the walk lies below $\delta_c$, is given by

$$P(s_n) = \int_{-\infty}^{\delta_c} \frac{d\delta_1}{\sqrt{2\pi}} \cdots \int_{-\infty}^{\delta_n} \frac{d\delta_n}{\sqrt{2\pi}} \frac{1}{|\det(\mathbf{C})|^{1/2}} \exp \left[-\frac{1}{2} \delta^T \mathbf{C}^{-1} \delta \right]$$

$$= \int_{-\infty}^{\delta_c} d\delta_1 \cdots \int_{-\infty}^{\delta_n} d\delta_n$$

$$\times \int_{-\infty}^{\infty} \frac{d\lambda_1}{2\pi} \cdots \int_{-\infty}^{\infty} \frac{d\lambda_n}{2\pi} \exp \left[\lambda^T \delta - \frac{1}{2} \lambda^T \mathbf{C} \lambda \right].$$

(A1)

where the superscript T denotes a transpose of an $n$-dimensional vector. Rescaling the variables using $\delta_i \rightarrow x_i = \delta_i / \sqrt{\delta} \lambda_i \rightarrow y_i = \lambda_i / \sqrt{\delta}$, and defining the normalized covariance matrix $r_{ij} = C_{ij} / \sqrt{\delta}$ and scaled barrier heights $y_i = \delta_i / \sqrt{\delta}$, this survival probability becomes

$$P(s_n) = \int_{-\infty}^{y_1} dx_1 \cdots \int_{-\infty}^{y_n} dx_n$$

$$\times \int_{-\infty}^{\infty} \frac{dy_1}{2\pi} \cdots \int_{-\infty}^{\infty} \frac{dy_n}{2\pi} \exp \left[\mathbf{y}^T x - \frac{1}{2} \mathbf{y}^T \mathbf{r} \mathbf{y} \right].$$

(A2)

In this language, the completely correlated limit is equivalent to setting $r_{ij} = 1$ for all $(i,j)$, since this will result in the joint probability distribution being a univariate Gaussian in say $x_1$, multiplying a product of Dirac delta’s $\delta(x_1 - x_1)$. This will lead to precisely the survival probability discussed in Section 2.3.

This language also allows us to understand the Peacock–Heavens ansatz. One can easily check that the discretized survival probability in equation (14) follows by choosing a sequence of $N < n$ steps $\{s^{(j)}\}$ and setting $r_{ij}$ to the block-diagonal form

$$r = \begin{pmatrix}
0 & 0 & 0 \\
0 & r^{(1)} & 0 \\
0 & 0 & r^{(2)} \\
0 & 0 & 0
\end{pmatrix}$$

(A3)

with $N$ blocks $\{s^{(j)}\}$ of length $(\delta_{j - 1} - \delta_{j - 1})$, each corresponding to a set of completely correlated steps $s^{(j)}_{ij} = 1$ for all $(i,j)$ in the $j$th block. The Peacock–Heavens prescription tells us to choose the sequence $\{s^{(j)}\}$ to be logarithmically equi-spaced with $(\Delta s_i) \sim \Gamma d^{(j)}$. This clarifies why the Peacock–Heavens ansatz is only an approximation, since the true structure of $r_{ij}$ is different. For example, for Gaussian smoothing with $n = -1$, we have $r_{ij} = 2\sqrt{\delta} \Gamma / (s_i + s_j)$, which is of
course unity along the diagonal, but falls off gradually rather than in sharp jumps. Nevertheless, as the main text shows, the subsequent steps in the ansatz (namely, the continuum limit and the evaluation of $\Gamma$) lead to a remarkably accurate prescription for the first crossing distribution.

APPENDIX B: TOPHAT CORRELATION PARAMETER

The Peacock–Heavens correlation parameter $\Gamma$ for an arbitrary filter and power spectrum follows from equations (3.24) and (3.15b) of Bond et al. (1991). The case of the Gaussian filter and generic power spectra was discussed in the main text. For a generic filter and power-law power spectrum $P(k) \propto k^n$, $\Gamma$ is given by

$$\Gamma = 2\pi \ln 2 \times \frac{\left( \int_0^\infty dx x^{n+2} W(x)^2 \right)^{1/2} \left( \int_0^\infty dx x^{n+4} W'(x)^2 \right)^{1/2} \left( \int_0^\infty dx x^{n+3} W(x) W(x)^2 \right)^{-1}}{\left( \int_0^\infty dx x^{n+3} W(x) W(x)^2 \right)^{-1}}, \quad (B1)$$

where $W(x) = W(kR)$ is the smoothing filter, and a prime denotes a derivative with respect to the argument. For the tophat, we have

$$W(x) = \frac{3}{x} j_1(x), \quad W'(x) = \frac{3}{x} \left( j_0(x) - \frac{3}{x} j_1(x) \right), \quad (B2)$$

where $j_0(x) = \sin(x)/x$ and $j_1(x) = (\sin(x) - \cos(x))/x^2$ are spherical Bessel functions of the first kind. It is useful to define the integrals

$$I_{00} = \int_0^\infty dx x^{n+2} j_0(x)^2,$$

$$I_{01} = \int_0^\infty dx x^{n+1} j_0(x) j_1(x),$$

$$I_{11} = \int_0^\infty dx x^n j_1(x)^2, \quad (B3)$$

in terms of which we can write $\Gamma$ for the tophat filter as

$$\Gamma = 2\pi \ln 2 \frac{|I_{01} - 3I_{11}|}{\sqrt{I_{00} I_{11} - I_{01}^2}}. \quad (B4)$$

The integrals in equation (B3) do not converge for all values of $n$. In particular, we have

$$I_{00} = 2^{-n} \sin(n\pi/2) \Gamma(n-1)(n-1)/4, \quad -3 < n < -1,$$

$$I_{01} = -2^{-n} \sin(n\pi/2) \Gamma(n-1)(n+1)/2, \quad -3 < n < -1,$$

$$I_{11} = 2^{-n} \sin(n\pi/2) \Gamma(n-1)(n+1)/(n-3), \quad -3 < n < 1, \quad (B5)$$

where $\Gamma(z)$ is the Euler gamma function.

The Peacock–Heavens correlation parameter is therefore defined for $-3 < n < -1$, and can be simplified to give

$$\Gamma = 2\pi \ln 2 \sqrt{(n+1)(n+3)/(n-3)}, \quad -3 < n < -1. \quad (B6)$$

The tophat filter also leads to results slightly different from the Gaussian, for the conditioned walks discussed in Section 4, in particular for the form of $r$. For example, if we set $x = R/R_\odot \leq 1$ then, for the tophat filter, $r = \sqrt{\chi/S}$ is $S$ and $S$ in the $S$ and $S$-plane. If $n = -1$, then $r = [2\{2x(1+x^2) + (1-x^2)^2 \log[(1-x)(1+x)]/(4x^2)]$, so $S/S \to 4/3$ as $S/S \to 0$. And if $n = 0$, then the real-space tophat is just like a sharp $k$-space filter, so $S/S \to 1$. Thus, in general, the numerical coefficient is smaller than when the filter is Gaussian.

APPENDIX C: STOCHASTIC BARRIERS

Treating the problem of halo formation by a single deterministic barrier is at best a crude approximation. In triaxial collapse models, themselves crude approximations, the collapse barrier is a function of three variables, the joint distribution of which depends on scale (Sheth, Mo & Tormen 2001). In excursion set language, at each step $n$, one asks if $\delta_n > \delta_0(e_n, p_n)$ (recall that $n$ is monotonically related to the smoothing scale). Since $e_n$ and $p_n$ are random numbers (they are neither Gaussian distributed nor independent of $\delta_n$), one may think of $\delta_n$ as being a function of $n$, so the problem is now that of a random walk ($\delta_n$) crossing a barrier ($\delta_0$) whose height is stochastic function of $n$. If projected on to the $(\delta, S)$-plane, this translates to scatter in the value of the critical density required for collapse and $S$ (e.g., fig. B1 in Sheth & Tormen 2002).

Since the stochasticity of the barrier is related to that of $e_n$ and $p_n$, and these have variance proportional to that of $\delta_0$, one expects the variance of the barrier height to be proportional to $S$. If one further assumes that changes in the barrier height are uncorrelated with changes in $\delta$, and that the steps in the barrier height are drawn from a Gaussian distribution (even though steps in $e$ and $p$ are not), then, for uncorrelated steps, the net effect of the stochastic barrier is simply to rescale all variances: $S \to S/\sigma$ for some $\sigma < 1$. This is most easily seen by noting that now one asks if $\delta_n > \delta_n + b_n$, where $b_n$ is a Gaussian number with variance $b_n = D S_n$, for some $D > 0$.

Since $\delta$ and $B$ are both Gaussian distributed, requiring $\delta_n + b_n > \delta_0$ is the same as requiring $g_n$, a Gaussian number with variance $S_n = S_0(1 + D)$, exceed $\delta_n$. This rescaling of the variance means that the problem is the same as considering when the previous sum of $n$ Gaussian numbers first exceeds $\delta_/\sqrt{1 + D}$. This is attractive because just such a rescaling appears to be necessary to reconcile the shape of the first crossing distribution with the measured counts of haloes in simulations: e.g. Sheth & Tormen (1999) suggest that $D \approx 0.4$. However, the rescaling of the variance which is most naturally associated with triaxial collapse models, $1 + D$ as not large as the factor of $1/0.7$ that is needed (Sheth, Mo & Tormen 2001), if the fundamental ansatz of equation (4) is correct (we noted in Section 5 that figs 2 and 3 of Sheth et al. 2001 suggest it is not).

For correlated steps, the issue is more complicated, since one must now decide if the same smoothing process which correlates the steps in $\delta$ also correlates the steps in the barrier height (and also: if the steps in $\delta$ were previously independent of those in the barrier height, does the smoothing now correlate them?). In triaxial collapse models, this is indeed the case ($e$ and $p$ are defined in the same physical volume as $\delta$). Hence, the steps in the barrier are correlated in the same way as for $\delta$, so the net effect is again to simply rescale variances. In particular, the Peacock–Heavens correlation parameter $\Gamma$ is not modified.

On the other hand, if the stochasticity in the barrier height is either due to processes on another scale or due to processes which have a different correlation structure than $\delta$, then $\Gamma$ will be modified. For example, the steps in barrier height are uncorrelated even when those in $\delta$ are, then $\Gamma \to \Gamma(1 + D)$, where $D$ is related to the variance in the barrier height (i.e. it also determines the rescaling of $S$). This is exactly analogous to the rescaling $x \to x(1 + D)$ which Maggiore & Riotta assume in their treatment of stochasticity. In a paper which appeared after an earlier version of the present work, Corasaniti & Achitouv (2011b) use this argument in the treatment of random walks and proceed to fit the resulting first crossing distribution to a mass function from $N$-body simulations using equation (4), leaving $D$ (and the barrier slope) as free
parameters and finding agreement at the 5 per cent level. We emphasize, however, that this rescaling is effectively invoking some unspecified process which operates on a scale that is not the same as that on which \( \delta \) was defined. More importantly, since the ansatz in equation (4) is most likely incorrect (see above), conclusions about how reasonable the fitted value of \( D \) is are suspect, given the obvious differences between figs 2 and 3 in Sheth, Mo & Tormen (2001).

And, finally, this treatment is motivated by the observed scatter in the value of \( \delta_c \) rather than on theoretical considerations of triaxial collapse – which introduces a new ambiguity as we discuss next.

C1 Distribution of deterministic barriers

The above discussion noted that if the collapse barrier is a stochastic function of scale, then this will appear as scatter in the critical density at fixed \( S \). However, it is important to note that the converse is not true: the observation of scatter in the critical density at fixed \( S \) (e.g. figs 2 and 3 of Sheth et al. 2001) does not imply that the barrier itself is stochastic. For example, one might find scatter in \( \delta_c \) at fixed \( S \) if there were simply a distribution of barrier shapes, each one of which was deterministic. To illustrate this (trivial) point, we consider three simple examples below. In all cases we illustrate our arguments using walks with uncorrelated steps, but note that the extension to correlated steps presents no conceptual difference.

C1.1 Constant barriers

The first assumes that the first crossing problem of interest is always that of a constant barrier; however, the barrier height is different for different walks. In this case, the distribution of interest is simply

\[
S_f(S) = \int d\delta_c p(\delta_c) S_f(S|\delta_c).
\]

If we use

\[
p(\delta_c) d\delta_c = \frac{4\pi \delta_c^2}{\Delta_c} \exp(-\delta_c^2/2\Delta_c^2) (2\pi)^{3/2}
\]

(C2)
to model the distribution of \( \delta_c \) around a typical value \( \Delta_c \), then the integral above can be done analytically:

\[
S_f(S) = \frac{2}{\pi} \frac{\Delta_c/\sqrt{\delta_c}}{(1 + \Delta_c^2/\delta_c^2)^{3/2}}.
\]

(C3)

This shows that the small \( S \ll \Delta_c \) tail is modified dramatically, from an exponential to a power law. If \( f(s) \) is indeed related to halo counts via equation (4), then the exponentially falling tail of halo counts argues against this as being the relevant model of stochasticity.

C1.2 Linear barriers

The model above has the same distribution of critical barrier heights at all \( S \). A simple model which allows \( S \)-dependent scatter is to assume that the deterministic barrier is linear, \( \delta_c + \beta S \), but that different walks have different values of \( \beta \). (This has some merit, since \( S \)-dependent scatter is suggested by figs 2 and 3 of Sheth et al. 2001.) If we use a Gaussian distribution for \( \beta \), then

\[
S_f(S) = \int d\beta p(\beta) S_f(S|\beta)
\]

\[
= \int d\beta \frac{e^{-\beta^2/2\Sigma_\beta^2}}{\sqrt{2\pi} \Sigma_\beta} \delta_c \frac{e^{-\delta_c^2/2S}}{\sqrt{2\pi}}
\]

\[
= \frac{\Sigma_\beta \delta_c}{\sqrt{(1 + \Sigma_\beta^2) S}} \frac{e^{-\delta_c^2/2(S+1\Sigma_\beta^2)}}{\sqrt{2\pi}}.
\]

(C4)

When \( \Sigma_\beta = 0 \), then this reduces to the solution for a constant barrier. So the effect of stochasticity is almost like rescaling \( S \rightarrow S(1 + \Sigma_\beta^2) \). In this case, the exponential cut-off at small \( S \) is not modified, but the distribution at large \( S \) is changed dramatically. The first term in the penultimate expression may be thought of as a correction to the deterministic \( \beta = 0 \) case. This factor is \( \exp(\delta_c^2/\Sigma_\beta^2) \) when \( S \ll \Sigma_\beta^2 \), but it decreases rapidly at large \( S \). Again, if equation (4) is correct, then the enhanced counts at small \( S \) are encouraging, but \( \Sigma_\beta \) may have to be rather large for this to be a competitive model. Of course, one is free to explore other models for \( p(\beta) \). We will not do so here, because we feel we have made our main point – a distribution of deterministic barriers produces a distribution of critical densities at fixed \( S \), and this distribution affects the quantity which enters equation (4).

C1.3 Square-root barriers

Before closing our discussion, we think it is worth showing explicitly that, by choosing the barrier shapes and their distribution appropriately, one can obtain effective first crossing distributions which closely approximate that for a single deterministic barrier with rescaled height (or, alternatively, with rescaled variance). This happens to be true if the deterministic barriers have height \( \delta_c + \beta \sqrt{S} \), with a Gaussian distribution of \( \beta \). This case is tractable, because the first crossing distribution of a square-root barrier (of specified \( \beta \)) is known. Although the exact solution is complicated (Breiman 1966), it is quite well approximated by

\[
S_f(S|\beta) \approx \left( 1 + \frac{\beta \sqrt{S}}{4 \delta_c} \right) \frac{\delta_c}{\sqrt{S}} \frac{e^{-\delta_c^2/2S}}{\sqrt{2\pi}}.
\]

(Sheth & Tormen 2002; Moreno et al. 2009), so that

\[
S_f(S) = \int d\beta \frac{e^{-\beta^2/2\Sigma_\beta^2}}{\sqrt{2\pi} \Sigma_\beta} S_f(S|\beta)
\]

\[
= \left( 1 - \frac{\Sigma_\beta^2}{4(1 + \Sigma_\beta^2)} \right) \frac{\delta_c}{\sqrt{(1 + \Sigma_\beta^2) S}} \frac{e^{-\delta_c^2/2(S+1\Sigma_\beta^2)}}{\sqrt{2\pi}}.
\]

(C6)

This final expression is remarkable, because it shows that \( S_f(S) \) has the same shape as the first crossing distribution of a barrier of constant height: i.e. \( \beta = 0 \) and \( \delta_c/\sqrt{1 + \Sigma_\beta^2} \). The constant of proportionality differs from unity, presumably because the approximation we used for the crossing of a square-root barrier is not quite normalized. But even for \( \Sigma_\beta = 0.65 \), for which the effective height is the desired \( 0.7 \delta_c \), this normalization factor is only 0.925. Therefore, we have not bothered to repeat this exercise using Breiman’s exact expression for the first crossing distribution.
C1.4 The potential to test

We conclude that such models for the stochasticity are potentially rather interesting. Perhaps more importantly, they are testable. By following essentially the same steps taken by Sheth et al. (2001) to produce their figs 2 and 3, it is rather straightforward to test if the stochasticity in $\Delta_c$ at fixed mass is indeed due to a distribution of deterministic barriers. The same is not true if the barrier height were a stochastic function of scale. Until such tests are performed, we argue that the question of how to correctly describe the observed scatter in barrier heights at fixed mass remains an open one. We leave this to future work.

This paper has been typeset from a \TeX/LaTeX file prepared by the author.