

Finite-size effects on the statistics of extreme events in the BTW model

A. Garber^a and H. Kantz^b

Max Planck Institute for the Physics of Complex Systems, Nöthnitzer Str. 38, 01187 Dresden, Germany

Received 17 September 2008

Published online 24 December 2008 – © EDP Sciences, Società Italiana di Fisica, Springer-Verlag 2008

Abstract The BTW Abelian sandpile model is a prominent example of systems showing self-organised criticality (SOC) in the infinite size limit. We study finite-size effects with special focus on the statistics of extreme events, i.e., of particularly large avalanches. Not only the avalanche size probability distribution, but also the mutual independence of large avalanches in the critical state is affected by finite-size effects. Instead of a Poissonian recurrence time distribution, in the finite system we find a repulsion of extreme events that depends on the avalanche size and not on the respective probability. The dependence of these effects on the system size is investigated and some data collapse is found. Our results imply that SOC is an unsuitable mechanism for the explanation of extreme events which occur in clusters.

PACS. 89.75.-k Complex systems

1 Introduction

The study of extreme events has gained considerable interest in recent years [1,2]. One could even define the dynamics of a system as complex, if it is capable of generating large fluctuations either autonomously or in response to infinitesimal perturbations, so that complex dynamics and dynamics with extreme events are intimately linked. Real world phenomena are, e.g., earthquakes, extreme weather events, stock market crashes, or large fluctuations in hydrodynamic flows such as wind gusts.

Self-organised criticality (SOC) [3] is clearly one of several dynamical mechanisms generating extreme events. A power law distribution for the event magnitude implies that there is no “largest possible event” and (depending on the precise value of the exponent) not even a well defined mean event size. In a list of key issues concerning extreme events in a given system, the quest for knowledge about the largest possible event as well as the mean return time of events exceeding certain thresholds would be the most prominent ones. In addition, more details of their statistics would be of interest, in particular the full return time distribution, from which one can learn whether events usually appear as clusters or whether there is a typical time interval between two successive events. Indeed, as we will show in this paper, SOC-type mechanisms in finite systems imply the existence of a “recovery time” thus precluding clusterisation of extreme events.

In this paper we focus on the Abelian sandpile model [4,5] as a particular SOC model in order to study

such statistical issues. Extreme events are in this case avalanches exceeding a predefined large threshold. This is motivated by the fact that also in most real world phenomena we would call an event “extreme”, if it exceeds a certain threshold, usually related to damage, casualties, or financial losses. Whereas for infinite system size the sandpile system is critical and all results concerning extreme events can easily be derived, we are particularly interested in finite-size effects. Again, the motivation stems from the fact that real world phenomena occur in systems of large but finite size, with a finite energy content, such that indeed it is conceivable that, e.g., magnitudes are restricted, so that there might exist the strongest possible earthquake or the highest possible wind speed in a storm.

In the following sections we will study the avalanche statistics of the sandpile model as a function of the system size. Our most relevant result will be a strong suppression of short recurrence times of large avalanches. We will show that this suppression is not related to a cutoff of the power law for the magnitude distribution, but instead to the inhibitive interaction of large successive avalanches in a finite system. Therefore, the finiteness of the system does not only cause a finite mean avalanche size and a finite maximal avalanche size, but also a non-Poissonian recurrence statistics with a typical timescale for the return of extreme events.

2 Avalanche size distribution

For our analysis, we use the Bak-Tang-Wiesenfeld model, which is a two-dimensional Abelian sandpile model defined

^a e-mail: garber@pks.mpg.de

^b e-mail: kantz@pks.mpg.de

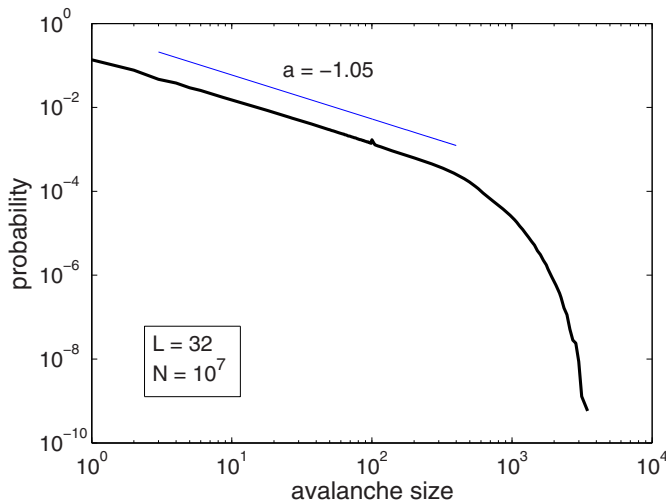


Figure 1. Probability distribution $p(s)$ of the avalanche sizes in the BTW model of system size $L = 32$, shown in log-log scale. A total of $N = 10^7$ simulated avalanches were used. The glitch in the data at $s = 100$ was introduced to mark the onset of logarithmic binning used to smooth the histogram. The straight line shows the power law.

on a square lattice with L^2 sites [4,5]. For each site, an integer variable z denotes the number of sand grains. Single grains are added to randomly chosen sites of the system. If the number of grains at a site (x, y) exceeds $z_c = 3$, the site will topple and distribute sand grains to the four nearest neighbours according to the following rules:

$$\begin{cases} z(x, y) \rightarrow z(x, y) - 4 \\ z(x \pm 1, y) \rightarrow z(x \pm 1, y) + 1 \\ z(x, y \pm 1) \rightarrow z(x, y \pm 1) + 1. \end{cases} \quad (1)$$

The boundary is open and dissipative, i.e. if (x, y) is a boundary site with only three or two nearest neighbours, one or two grains will leave the system. If as a result of the added sand grain a neighbour becomes unstable, it also topples, starting an avalanche during which no new sand grains are added to the system. The avalanche stops when all sites are stable again, its size is defined as the total number of topplings that have taken place. Contrary to the recommendation made by Dhar to minimise finite-size effects [6], the avalanches starting near the boundary were not omitted from the analysis.

As can be seen from Figure 1, the probability distribution $p(s)$ of avalanche sizes does indeed obey a power law, whose exponent has been estimated by a number of authors as lying between -1.05 and -1.293 depending on the system size used in the simulations [7–10]. For large avalanche sizes however, the finite system size leads to a truncation of the power law. This cutoff appears at avalanche sizes of approximately $L^2/2$.

The truncation also leads to the existence of a mean avalanche size. Our simulations show that this mean scales with $L^{1.9}$. Using this result, the scaling of the cross-over to the truncation regime can be verified by a rough approximation. Setting the avalanche size probability to zero for

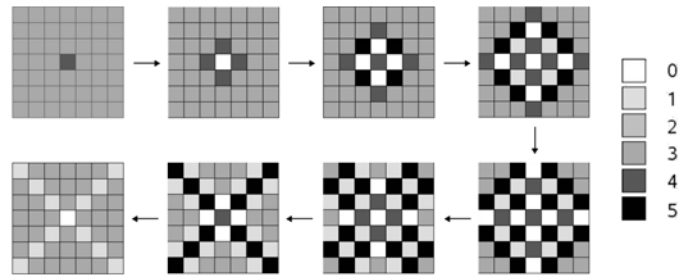


Figure 2. Spread of the maximal avalanche for $L = 7$. The height of sand grains at each system site is colour coded. Each square represents one sweep in the toppling algorithm (during a sweep all unstable sites topple simultaneously).

sizes outside the power law regime, the mean avalanche size can be calculated as

$$\langle s \rangle \simeq \int_0^{\frac{L^2}{2}} s \cdot p(s) ds \propto L^{1.8} \quad \text{for } p(s) \propto s^{-1.1}. \quad (2)$$

Considering the crudeness of this approximation, its result coincides nicely with the empirical result. This demonstrates the mutual consistence of the empirical scaling laws for $p(s)$, the cut-off of the power law, and the mean avalanche size.

Another consequence of the dissipation of sand grains over the border of the system, which cannot directly be seen from the probability distribution of avalanche sizes, is the existence of a maximal avalanche size which can be calculated analytically. For the maximal avalanche, the probability of a toppling causing other sites to become unstable has to be maximised. This is achieved by a uniform starting distribution of three grains on each site – the maximum, while still retaining the stable state of the system – with a single extra grain placed in the middle of the system to start the avalanche, ensuring that the first sand grain leaves the system at the latest possible moment. Figure 2 shows the evolution of the resulting avalanche dynamics from the initial configuration, splitted into sweeps of simultaneously unstable system sites. Two phases of the avalanche become visible. At first, the spread of toppled sites is small enough not to reach the border of the system so the number of topplings per sweep increases as a sequence of squares. When the spread of the avalanche reaches the border, the number of sites in this sequence of squares that lie outside the border as well as the number of sites which do not become unstable again due to a lack of sand grains re-entering the system from outside have to be subtracted. Summing up these contributions, we arrive at the following formula for the maximal avalanche in a system of size L (L odd):

$$\begin{aligned} s_{max} &= \sum_{i=1}^L i^2 - 4 \sum_{i=1}^{\frac{L-1}{2}} \sum_{k=1}^i k - 4 \sum_{i=1}^{\frac{L-3}{2}} \sum_{k=1}^i k \\ &= \frac{1}{6} L(L+1)(L+2). \end{aligned} \quad (3)$$

For even system sizes, the single extra grain is placed at one of the middle sites of the system, resulting in a slightly

different calculation of the maximal avalanche because of the lower degree of symmetry. The result, however, is the same as for odd system sizes, equation (3).

The existence of characteristic avalanche sizes such as the maximum and the mean shows that the influence of the finite system size has a major impact on the scale-free behaviour of avalanches in the BTW model. The system size hence introduces scales into the statistics of extreme events.

3 Extreme value statistics

The mathematical theory of extreme value statistics considers the maxima of blocks of n successive events. Under the assumption that successive events are independent of each other, the distribution of block maxima follows the universal Generalised Extreme Value distribution (GEV) as described by equation (4), which comprises the Fréchet ($\xi > 0$), Gumbel ($\xi = 0$) and Weibull ($\xi < 0$) distributions distinguished by the shape parameter ξ ,

$$G(z) = \exp \left\{ - \left[1 + \xi \left(\frac{z - \mu}{\sigma} \right) \right]^{-\frac{1}{\xi}} \right\}. \quad (4)$$

The type of extreme value statistics itself is used for the extrapolation from the statistics of observed event magnitudes during past observation periods to the statistics of unobserved event magnitudes and future, much longer observation periods. Even though this theory is only valid for independent events with a stationary distribution, rapidly decaying correlations are not relevant for big block sizes n [11].

Avalanches in the Abelian sandpile model with infinite system size are power law distributed. Evaluating the criteria for the domain of attraction of each of the GEV families [12], their block maxima should be Fréchet distributed. However, as shown in Section 2, the finite system size introduces both a truncation of the power law and a maximal avalanche. The expected distribution of the block maxima in the presence of such an upper bound would then be of the Weibull family.

Exemplary block maxima distributions for $L = 64$ and two different block sizes n are displayed in Figure 3. As can be seen, for the block size $n = 100$, when many smaller avalanches still within the power law regime are considered, the shape parameter indicates the expected Fréchet distribution, while for the larger block size $n = 8000$, the maxima instead follow a Weibull distribution. This transition is better visible in Figure 4, where the dependence of the shape parameter ξ on the block size n is shown.

The GEV distributions are expected under the condition of the independence of successive avalanches and the stationarity of their distribution. If these conditions are violated, other distributions might be found. Our fits demonstrate that indeed our empirical distribution is well described by the members of the GEV. This is not unexpected. Since stationarity is a prerequisite of the non-equilibrium steady state of SOC, this is not a concern. The

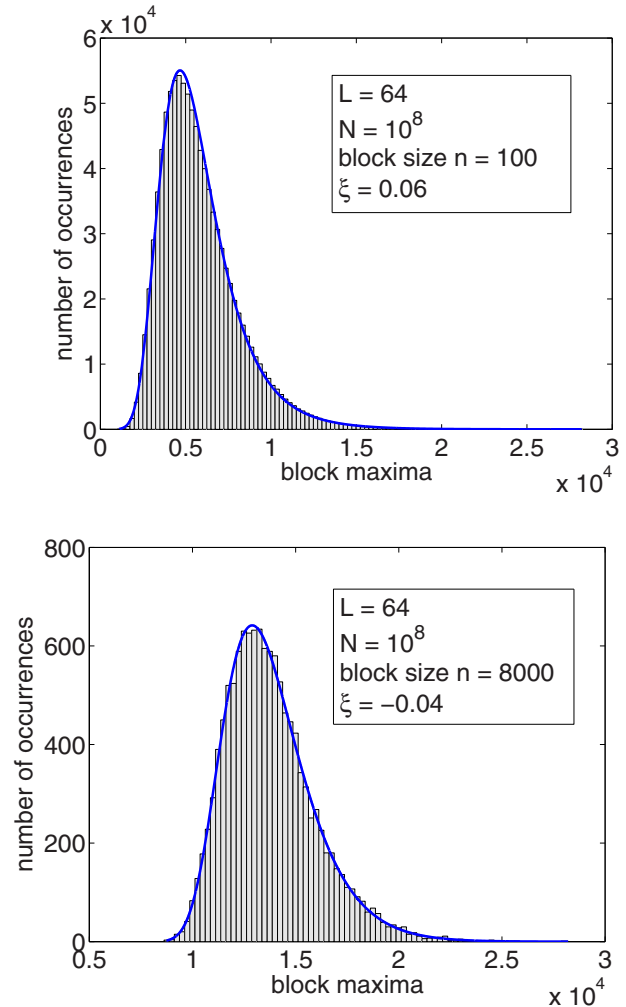


Figure 3. Histograms of the block maxima for $L = 64$ and block sizes $n = 100$ (upper) and $n = 8000$ (lower) with the corresponding GEV distribution (smooth curves).

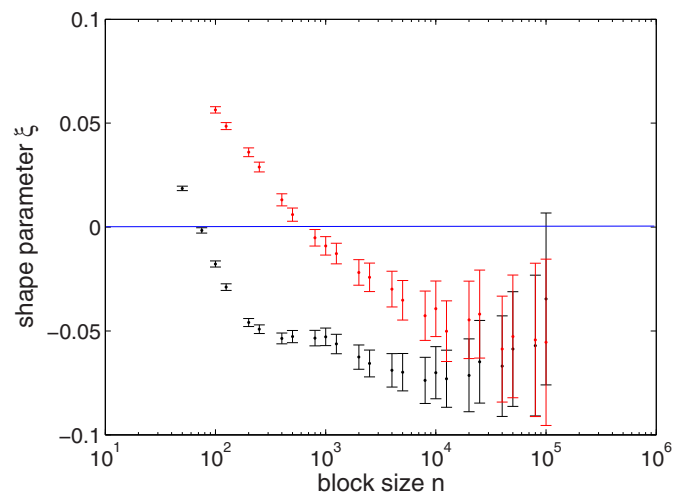


Figure 4. Dependence of the shape parameter ξ on the block size n in a semilogarithmic scale for the system sizes $L = 16$ (lower family of values) and $L = 64$ (upper). The error bars represent the 95% confidence level.

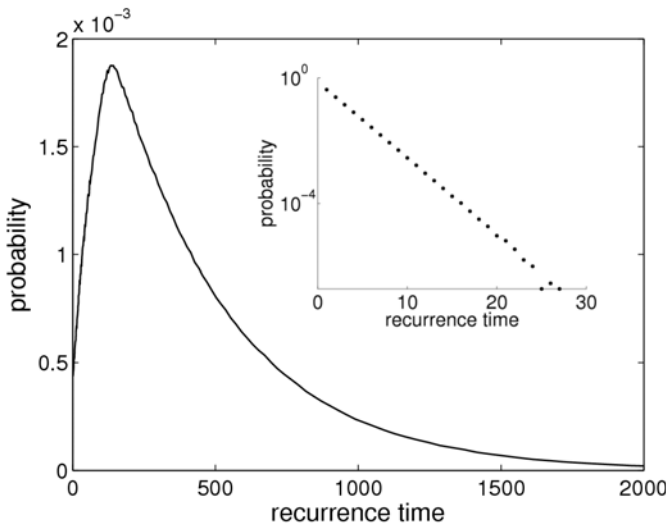


Figure 5. Probability distribution of the recurrence time of avalanches of size $s \geq s_0 = 1120$ for $L = 32$ and $N = 10^9$ simulated avalanches. For clarity the recurrence time axis was truncated at $t = 2000$ but the largest recurrence time occurring in the simulation was $t = 6279$. The inset shows the same probability distribution for $s_0 = 1$ in a semilogarithmic scale.

analysis in Section 4 shows that small avalanches follow a Poissonian recurrence time distribution and are indeed independent. While larger avalanches are correlated, these correlations are shorter than the large block sizes used for the extreme value statistics and are therefore not relevant, justifying the application of this theory.

4 Recurrence time

In an infinite system, the probability that two consecutive avalanches start close enough to influence each other is very small. Therefore it is very plausible that in this case the avalanches can be considered as independent events. Other arguments in favour of independence have been published [13] and verified by simulation [14]. The recurrence time of avalanches equal to or exceeding a given threshold s_0 , measured as the number of added sand grains between their occurrence, is therefore expected to be distributed exponentially as $p(t) = \tau^{-1} \exp(-t/\tau)$, where τ is the mean recurrence time which depends on the avalanche size under study.

Analysing the recurrence time of avalanches without a threshold s_0 , the Poissonian distribution is indeed observed, as shown in the inset of Figure 5. However, for larger thresholds s_0 , finite-size effects appear. Figure 5 shows the probability distribution of recurrence times for $L = 32$ and for large s_0 . It accurately follows the expected exponential distribution for large enough times t , but the probability of shorter times is significantly reduced. This implies that the system becomes undercritical by an extreme event and needs to build itself up to the critical state before avalanches of a similar size can again occur,

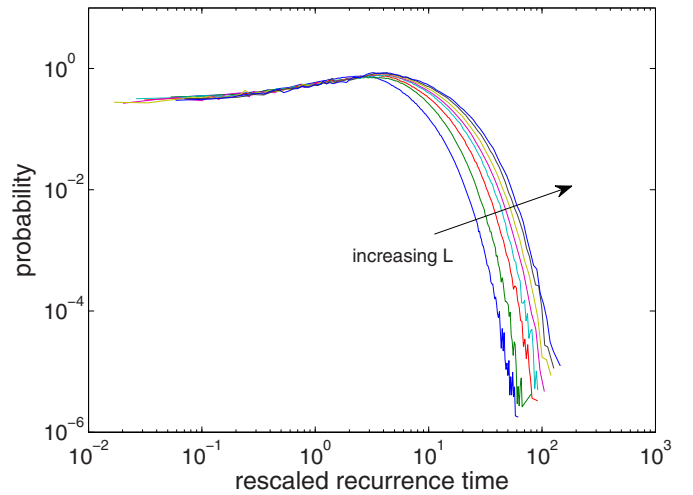


Figure 6. Data collapse of recurrence times of large avalanches in systems of different sizes equally spaced between $L = 16$ to $L = 72$ (log-log scale was used to emphasize small times). The curves have been shifted horizontally by a factor of $L^{0.6}$. A simulation of a total number of $N = 10^8$ avalanches was used.

showing that the finite system size not only prevents total scale-freeness, but also causes a correlation between extreme avalanches.

In order to investigate the dependence of these correlations on the system size and also on the definition of extreme avalanches, we search for a rescaling such that a data collapse is achieved. Implementing Kac's Lemma [15] it is possible to calculate the mean recurrence time $\tau = N/n_{large}$, where N is the total number of added sand grains and n_{large} is the number of avalanches equal to or exceeding the chosen threshold s_0 . Measuring the recurrence time in units of $\tau(L)$, the exponential part of the distributions of recurrence times for different system sizes L collapses rather well. Any remaining deviations are due to the normalisation constraint that spreads the influence of finite-size effects beyond small recurrence times.

The shape of the finite-size effect for small recurrence times is determined by the threshold s_0 defining extreme avalanches. The larger s_0 , the stronger the depletion around $t = 1$ is. For a data collapse of the relevant part of the probability distributions, appropriate thresholds $s_0(L)$ have to be chosen. This is not quite enough for a full data collapse, the data also has to be shifted along a horizontal direction. This does not maintain the normalisation constraints, but no valuable information is lost in the process as the original curves can be easily restored.

An example of the resulting collapsed probability distributions can be seen in Figure 6, where one such sequence $s_0(L)$ has been determined empirically, which under rescaling collapses in the region of small t . To determine this sequence, first, a value for $s_0(L = 16)$ has been chosen. Then, for $L \neq 16$, threshold values have been varied until values were found such that the shifted curve collapses onto the curve for $L = 16$. By this, a series of

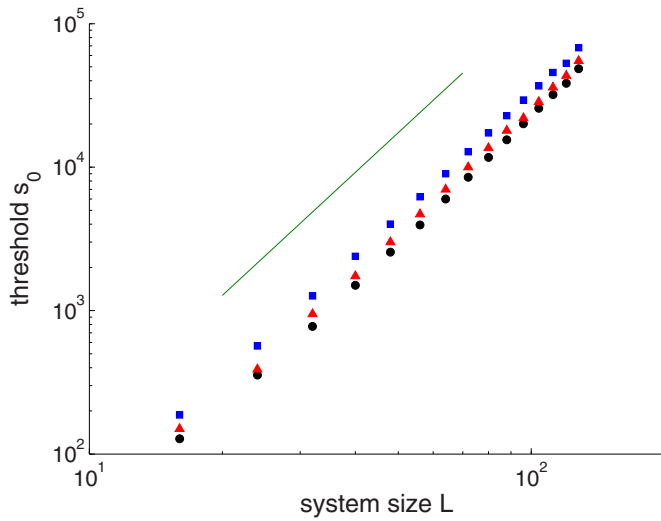


Figure 7. Three different series of thresholds leading to different strengths of the finite-size effect in large avalanches and their dependence on the system size L in a log-log-scale. The uncertainty of the vertical position is of the order of the symbol size. The straight line illustrates a power law of exponent 2.87.

thresholds $s_0(L)$ has been determined which all collapse onto a single curve in the range of short recurrence times.

Figure 7 shows three series of thresholds $s_0(L)$ for system sizes $L = 16$ to $L = 128$ corresponding to three distinct shapes of the finite-size effect on the probability distribution of recurrence times. These series of thresholds all depend on the system size as a power law with an exponent of $\alpha \approx 2.87 \pm 0.07$. The error stems from the estimation of the determination error of the thresholds $s_0(L)$.

5 Interdependence

The results until now being purely empirical, a more thorough understanding of the mechanisms leading to the observed correlations between large avalanches is needed, specifically of the parameter dependence $s_0(L)$ leading to the observed power law of Figure 7.

The first hypothesis about how the threshold values $s_0(L)$ are controlled by L is based upon the avalanche probabilities. It might be that independently of L the thresholds $s_0(L)$ lead to the same event rate, i.e. to the same probability for an avalanche to exceed the threshold. In order to check this, the mean recurrence times of avalanches exceeding the thresholds $s_0(L)$ are plotted versus the system sizes (Fig. 8). As can be seen, the dependence is nontrivial, so thresholds for different system sizes that lead to the same probability of an extreme avalanche occurring do not lead to the same shape of the finite-size effect.

If, on the other hand, the thresholds $s_0(L)$ were determined in such a way that the probability of an equal-to-threshold avalanche were the same for all system sizes L ,

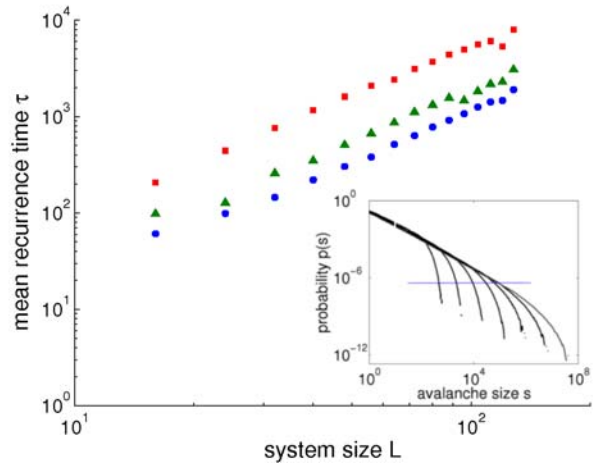


Figure 8. Mean recurrence time for the three different series $s_0(L)$ in a log-log scale. The inset shows the avalanche size probability distribution for different system sizes L ranging from $L = 16$ to $L = 1024$.

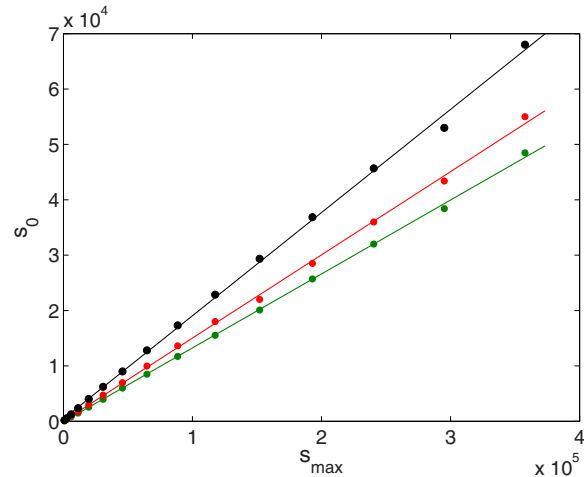


Figure 9. Dependence of the series of thresholds $s_0(L)$ on the maximum avalanche size $s_{max}(L)$.

then the series would be given by a set of s -values determined by the intersection of the horizontal line in the inset of Figure 8 with the probability distributions. We see that this is not plausible, since for small systems the threshold would be far in the cutoff regime beyond the power law, whereas for large L it were inside the power law regime, where no finite-size effects are to be expected.

Comparing the empirical power law scaling $s_0(L) \approx L^{2.87 \pm 0.07}$, and the numerical fit to equation (3) for our range of system sizes, $s_{max} \approx L^{2.93}$, one might also surmise that the thresholds all lead to a consideration of only those avalanches whose size surpasses a given percentage of the maximal avalanche. Figure 9 shows indeed a good agreement with a linear dependence of the threshold s_0 on the maximum avalanche s_{max} despite deviations for the smaller system sizes. Therefore, a simple criterion for the determination of the thresholds was found, but the detailed mechanisms leading to the precise series $s_0(L)$ still remain unclear.

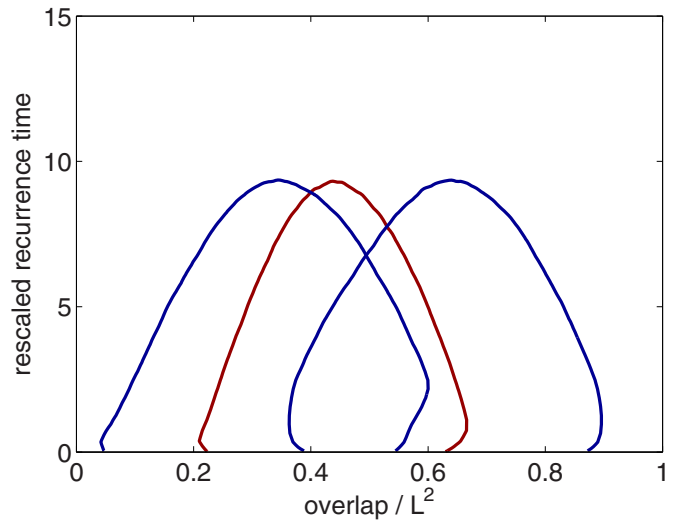
Hence, a more involved reasoning is needed. The suppression of short recurrence times means that large avalanches repel each other. This repelling mechanism might have its origin in the spatial overlap between two consecutive avalanches that exceed the chosen threshold s_0 , defined as the number of system sites toppling during both avalanches. We determined this quantity numerically. It obeys an approximately Gaussian statistics, but the mean overlap increases with system size for each series of $s_0(L)$, implying that the influence of the overlap does not seem to be the determining factor even after normalisation to the total number of system sites.

On the other hand, a 2D probability distribution of overlap and recurrence time could be better suited to investigate a possible dependence. If those two quantities were uncorrelated, this distribution would be the product of a Gaussian in horizontal direction and the modified exponential distribution in vertical direction with the observed finite-size suppression of small recurrence times. The right contour in Figure 10a shows this direct product of the one-dimensional distributions. For comparison, the other two contours show the observed empirical distributions for different parameter values s_0 . The finite-size deviation for small recurrence times is visible for each contour, but the symmetry of the Gaussian distribution is violated as the deviation is significantly increased for large values of the overlap. As can be seen from Figure 10a, this increase depends on the value of s_0 . Figure 10b indicates that a similar correlation between large overlaps and small recurrence times is a selection criterion behind the thresholds s_0 leading to the same shape of the finite-size effect on recurrence times.

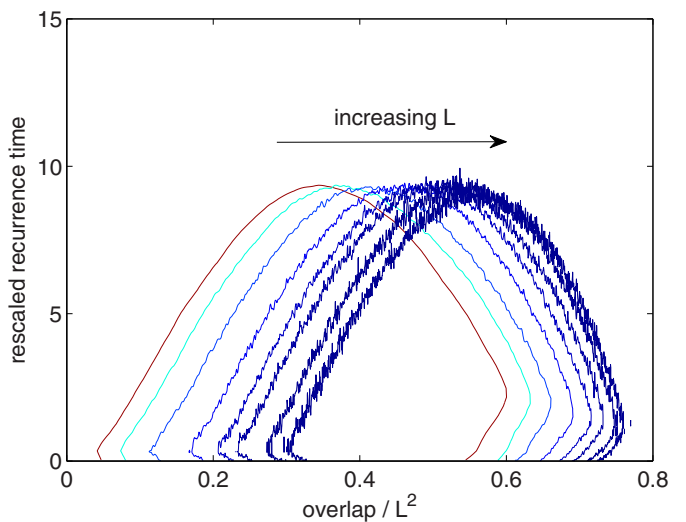
Interestingly, in Figure 10a, the bigger the avalanches considered, the smaller the depletion of recurrence times for a large overlap is. Since large avalanches should be more influenced by the finite system size, this seems at first counter-intuitive. However, the analysis of the overlap disregards avalanches of size $s < s_0$ occurring between two avalanches exceeding this threshold. These smaller events redistribute grains of sand in the system thus connecting avalanches without any actual overlap and decorrelating those with a large overlap. The higher the threshold s_0 , the larger the mean recurrence time, allowing more small events to take place in between. This interrelation between overlap and mean recurrence time makes it difficult to gain a more thorough understanding of the mechanisms leading to the empirical series $s_0(L)$.

6 Conclusion

We have shown that the finite size of the BTW sandpile has a strong influence on the key properties of extreme events. The well known violation of the scale invariance of avalanche sizes is relevant in this context since it ensures the existence of both a mean and a maximal avalanche size that scale with the system size as power laws. In addition, large avalanches are not independent but repel each other, resulting in a significantly lowered probability of small recurrence times depending on the chosen threshold for large



(a) Contours for $L = 16$ and – from left to right – $s_0 = 128$, $s_0 = 187$, and the direct product of the one-dimensional distributions for $s_0 = 128$ (shifted by $+0.3$ in the horizontal direction for clarity).



(b) Contours for a series of thresholds $s_0(L)$ leading to the same magnitude of the finite-size effect for different L .

Figure 10. Contour plots of the 2D probability distribution of overlap and recurrence time.

avalanches. For very large thresholds, the recurrence time distribution is strongly peaked, introducing a typical time interval in between two successive extreme events.

Several different hypotheses about the mechanism determining the shape of this finite-size effect were tested. We found that the thresholds do not lead to the same probability of obtaining a large avalanche exceeding the threshold. However, within the accuracy of the simulation and analysis, the thresholds are at a given fraction of the maximal avalanche size. The overlap between avalanches exceeding the threshold may also play a role, but as two above-threshold avalanches do also interact through

smaller sub-threshold avalanches, this cannot be analysed in a meaningful way.

The non-Poissonian return time statistics found in this SOC model causes a more homogeneous distribution of extreme events over time. It strongly suppresses clusters of extreme events. This is in clear contrast to systems exhibiting long range correlations [16], where the return time statistics was found to be a stretched exponential or Weibull distribution with the enhancement of clustering of events. Thus, the return time statistics might be used to infer (or at least exclude) certain dynamical mechanisms for the creation of extreme events. More precisely, we conclude that phenomena where extreme events tend to form clusters in time cannot be caused by SOC-type mechanisms.

References

1. *The Science of Disasters, Climate Disruptions, Heart Attacks, and Market Crashes*, edited by A. Bunde, J. Kropp, H.J. Schellnhuber (Springer, Berlin, 2002)
2. *Extreme Events in Nature and Society*, edited by S.A. Albeverio, V. Jentsch, H. Kantz (Springer, Berlin, 2006)
3. P. Bak, C. Tang, K. Wiesenfeld, Phys. Rev. Lett. **59**, 381 (1987)
4. P. Bak, C. Tang, K. Wiesenfeld, Phys. Rev. A **38**, 364 (1988)
5. D. Dhar, Phys. Rev. Lett. **64**, 1613 (1990)
6. D. Dhar, Physica A **369**, 29 (2006)
7. S. Lübeck, K.D. Usadel, Phys. Rev. E **55**, 4095 (1997)
8. L. Kadanoff, S. Nagel, L. Wu, S. Zhou, Phys. Rev. A **39**, 6524 (1989)
9. S. Manna, Physica A **179**, 249 (1991)
10. E. Milshtein, O. Biham, S. Solomon, Phys. Rev. E **58**, 303 (1998)
11. S. Coles, *An Introduction to Statistical Modeling of Extreme Values* (Springer-Verlag, London, 2001)
12. L. de Haan, A. Ferreira, *Extreme Value Theory, An Introduction* (Springer, New York, 2006)
13. G. Boffetta, V. Carbone, P. Giuliani, P. Veltri, A. Vulpiani, Phys. Rev. Lett. **83**, 4662 (1999)
14. M. Paczuski, S. Boettcher, M. Baiesi, Phys. Rev. Lett. **95**, 181102 (2005)
15. M. Kac, Bull. Am. Math. Soc. **53**, 1002 (1947)
16. A. Bunde, J.F. Eichner, J.W. Kantelhardt, S. Havlin, Phys. Rev. Lett. **94**, 048701 (2005)