PHYSICAL REVIEW E

STATISTICAL PHYSICS, PLASMAS, FLUIDS, AND RELATED INTERDISCIPLINARY TOPICS

THIRD SERIES, VOLUME 54, NUMBER 2 AUGUST 1996

RAPID COMMUNICATIONS

The Rapid Communications section is intended for the accelerated publication of important new results. Since manuscripts submitted to this section are given priority treatment both in the editorial office and in production, authors should explain in their submittal letter why the work justifies this special handling. A Rapid Communication should be no longer than 4 printed pages and must be accompanied by an abstract. Page proofs are sent to authors.

Word processors with line wrap: Cascading, self-organized criticality, random walks, diffusion, and predictability

Wolfgang Bauer and Scott Pratt

Department of Physics and Astronomy and National Superconducting Cyclotron Laboratory, Michigan State University, East Lansing, Michigan 48824-1321

(Received 20 May 1996)

We examine the line-wrap feature of text processors. We show that adding characters to previously formatted lines leads to the cascading of words to subsequent lines. The length of these cascades shows a power-law distribution. We show that this system is in a state of self-organized criticality. The connection to one-dimensional random walks and diffusion problems is demonstrated. Of particular interest is the exponential cutoff of the power-law distribution occurring for finite line lengths. Finally we examine the predictability of large cascades. [S1063-651X(96)50208-4]

PACS number(s): 05.40.+j, 07.05.Tp, 89.80.+h, 91.30.Px

Composite systems may evolve to a critical state in which minor events may trigger a chain reaction that can affect an arbitrarily large number of constituents of the system. This state was called *self-organized criticality* [1–3], and was first investigated for sandpiles. Theoretically and experimentally [4], avalanches in sandpiles show power-law distributions characteristic of real earthquakes [5,6]. The absence of an intrinsic length scale is attributed to self-organized criticality, where avalanches of all sizes contribute to keep the system perpetually in a critical state.

So far, to our knowledge, no analytic solution to the model of Ref. [1] has been presented. However, if one introduces a preferred direction, then an exact analytic solution is possible [7], and a connection to the problem of two annihilating random walkers can be established. Other authors have established connections to the driven or convective diffusion equation [8]. A number of one-dimensional (1D) sandpile models with nontrivial dynamics have been proposed [9], some of which can be analytically solved. Other interesting examples of self-organized criticality have been found in models of forest fires [10] and evolution [11].

Of particular interest is the question of predictability of catastrophic avalanches. In a recent experiment measuring the total mass of the sandpile it was found that a running total of small avalanches can predict the occurrence of large avalanches [12].

Here we introduce what we believe to be the most simple (and most directly connected to everyday experiences) example of self-organized criticality. Consider a modern word processor with line-wrap feature and fixed maximum number of characters per line. Such a word processor formats a paragraph without the explicit need to enter carriage returns or line feed characters. If a word is too long for a line, it is automatically wrapped into the next line. We consider an infinitely long paragraph formatted by this word processor. If one then adds another character to the beginning of the first line, the last word of this line may be wrapped to the second line, and a cascade of line wraps may ensue. If a steady stream of characters or words is entered at the beginning of the first line, a sequence of cascades on all scales results, and the line lengths (excluding trailing blanks) in the paragraph form a self-organized critical state. In the following we will show that finite line lengths set a length scale in this system which cuts off the scale-invariant behavior. We will also show the connection to random walks and the diffusion equation. And finally we will address the important question of predictability of large avalanches in this system.

Let us—for the moment—neglect temporal correlations between the individual words shifted through the paragraph.

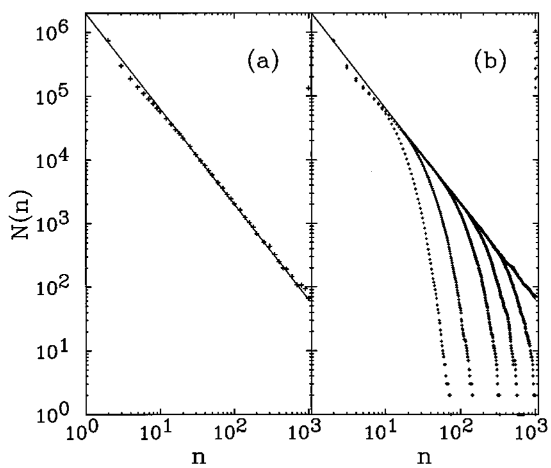

FIG. 1. Cascade lengths distribution for the word processor with line-wrap feature. (a) Infinite line length, (b) finite line lengths, $\ell_l = 20$, 30, 50, 70, 100, 200, from left to right, respectively. $N(n)$ denotes the number of times a cascade occurred that affected exactly the first n lines. The straight lines correspond to the asymptotic $n^{-3/2}$ solutions.

(This assumes in practice that individual lines are infinitely long—an assumption we will relax again below.) In addition, we assume here for definiteness that the individual word lengths including one trailing blank are evenly distributed between 2 and $2\langle \ell_w \rangle - 2$, where $\langle \ell_w \rangle$ is the average word length. [The probability for any given word length between these two limiting values is then $(2\langle \ell_w \rangle - 3)^{-1}$.] For a fixed number of characters per line, the probability to have b trailing blanks in this line can be shown to be

$$p(b) = \begin{cases} \langle \ell_w \rangle^{-1} & \text{for } b = 0,1 \\ \dfrac{2\langle \ell_w \rangle - 2 - b}{2\langle \ell_w \rangle^2 - 3\langle \ell_w \rangle} & \text{for } 2 \leqslant b \leqslant 2\langle \ell_w \rangle - 2 \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

and consequently the average number of trailing blanks in a line is

$$\langle b \rangle = \sum_{b=0}^{2\langle \ell_w \rangle - 2} b p(b) = \frac{2}{3}\langle \ell_w \rangle - 1 + \frac{1}{3}\langle \ell_w \rangle^{-1}. \quad (2)$$

In Fig. 1(a) we display (crosses) the size distribution for cascades in this system. $N(n)$ denotes the number of times a cascade occurred that affected exactly the first n lines. Here the total number of lines was chosen to be 10^3, and $\langle \ell_w \rangle = 6$. A total of 3×10^6 words ($= 1.8 \times 10^7$ characters) were generated for this plot. One can clearly observe that for large number of lines, n, the distribution approaches the power-law limit (solid line),

$$N(n) \propto n^{-3/2}. \quad (3)$$

It is also of interest to compute the total activity, i.e., the total number of characters moved to different lines. The number distribution for this activity also approaches a power law, with an exponent of approximately $-4/3$ very much reminiscent of the earthquake strength distribution found in Ref. [5].

We should point out here that none of our results depends on the type of the word length distribution chosen. We obtained for practical purposes identical results with Poissonian word length distributions.

The result of Eq. (3) can be understood by formulating the problem in terms of a random walk. One step in this random walk is the change in the number of blanks in a given line caused by a cascade passing through. To derive the step size distribution, we realize that pushing a word of length ℓ from line n to $n+1$ increases the number of trailing blanks in line n by ℓ. Conversely, pushing a different word of length ℓ' from line $n-1$ into line n decreases the number of trailing blanks in line n by ℓ'. The probability distribution for a change Δb in the number of trailing blanks in a line is then

$$P(\Delta b) = \sum_{b=0}^{\infty} p(b) p(b - \Delta b) \quad (4)$$

where $p(b)$ is the probability distribution for trailing blanks as given by Eq. (1). The probability distribution $P(\Delta b)$ is symmetric about $\Delta b = 0$ and approximately triangular in shape. Its mean is 0, and its variance is finite.

The total number of characters moved through line n by the cascade is

$$c_n = -\sum_{i=1}^{n-1} \Delta b_i, \quad (5)$$

where Δb_i is the change of the number of trailing blanks in line i. If, for any n, we have $c_n \leqslant b_n$, where b_n = number of trailing blanks in line n, then the cascade terminates. Thus we see that our cascading problem is equivalent to a random walk problem with step size distribution given by Eq. (4). The result of Eq. (3) is the solution of the return-to-the-origin problem for a one-dimensional random walk.

It is, perhaps, more instructive to consider the corresponding continuum diffusion problem. The diffusion equation is

$$\partial_t f(x,t) = D \partial_x^2 f(x,t) \quad (6)$$

with the boundary condition $f(0,t) = 0$ and the solution

$$f(x,t) = \frac{x}{4\sqrt{\pi D^3 t^3}} \exp[-x^2/4Dt], \quad (7)$$

where t corresponds to the number of lines, n, in the random walk, and x is the distance of the random walk to 0. D is the diffusion constant and can be calculated from the second moment of the random walk step size distribution, Eq. (4),

$$D = \langle \Delta b_n^2 \rangle / 2. \quad (8)$$

($D \approx 6.3$ for the parameters used to produce Fig. 1.) For the current at the origin we obtain

$$J(t) = D \partial_x f(x,t)|_{x=0} \propto t^{-3/2}, \quad (9)$$

in agreement with the numerical finding of Fig. 1.

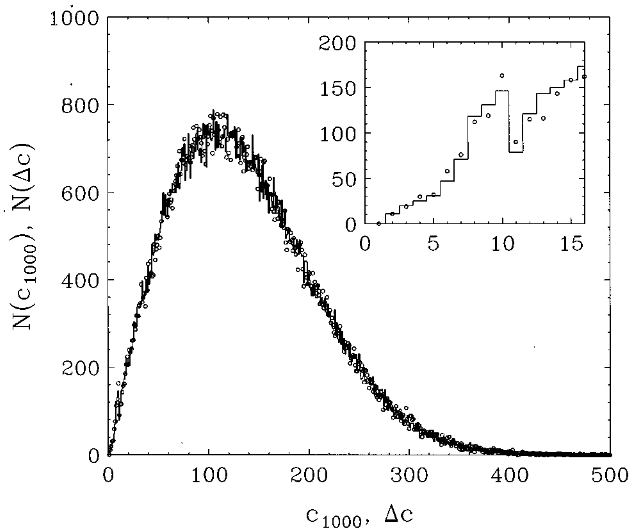

FIG. 2. Histogram: Number distribution, $N(\Delta c)$, of events with Δc characters entered between catastrophic cascades, i.e., cascades which reached line 1000. Circles: Number distribution, $N(c_{1000})$, of events with total length of random walks, c_{1000}, for catastrophic cascades. The inset is a magnified view of the region around the origin. The parameters of this simulation are identical to the ones used for Fig. 1(a).

We now proceed to study the case where we include all of the temporal correlations entailed by pushing an ordered (but individually randomly selected) sequence of words through our word processor. This is the case for finite line lengths. Inserting a number of characters equal to the line length, ℓ_l, will result in a completely new first line, pushing the old first line into the second, and so on. Thus we get catastrophic cascades (= cascades involving all lines—1000 in the specific example considered here) at least every ℓ_l characters.

In Fig. 1(b) we show our results using identical parameters to Fig. 1(a), but using finite line lengths, $\ell_l = 20$, 30, 50, 70, 100, and 200. It can clearly be seen that the power-law distribution is now cut off by an exponential. The exponent is numerically found to be $\propto(\ell_l)^{-2}$. This behavior can be understood in terms of the random walk formulation of the problem. The finite line length corresponds to an additional absorbing barrier for the random walk, restricting $c_n \leqslant \ell_l \;\forall n$. The corresponding solution to the diffusion equation is

$$f_f(x,t)=\sum_{j=1}^{\infty} C_j \sin(k_j x)\exp(-Dk_j^2 t) \tag{10}$$

with long-time behavior dominated by k_1, where $k_j = \pi j/\ell_l$. For large word lengths, we have $D \propto \langle \ell_w \rangle^2$ and the exponent thus is $\propto(\langle \ell_w \rangle/\ell_l)^2$.

Of particular interest in studying models with ''random'' catastrophic events is to investigate the limits of predictability of these events. To do this we record the number of characters, Δc, entered between catastrophic cascades. Using the same parameters as for the calculations in Fig. 1, we display in Fig. 2 (histogram) the number of events as a function of Δc, $N(\Delta c)$. It is obvious from this figure that they follow a Wigner distribution,

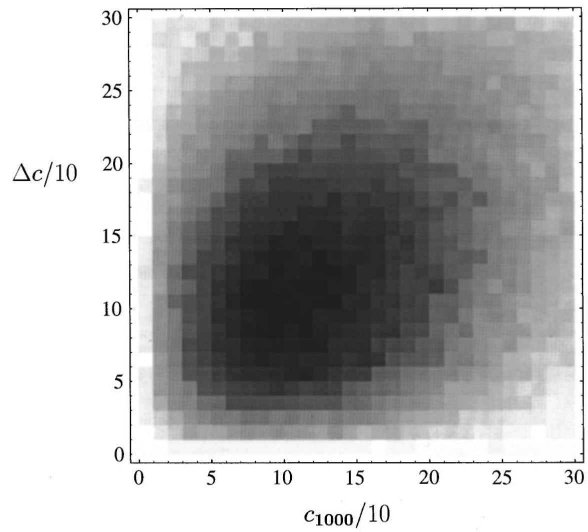

FIG. 3. Contour plot of the number distribution, $N(c_{1000},\Delta c)$, of catastrophic cascades with random walks of total length c_{1000} and number of characters entered, Δc, before the next catastrophic cascade. The gray level is proportional to $N(c_{1000},\Delta c)$, with black representing the maximum and white a value of 0.

$$N(\Delta c)\propto(\Delta c/\langle \Delta c\rangle)\exp[-\pi(\Delta c/\langle \Delta c\rangle)^2/4]. \tag{11}$$

(As a side note we mention here that we obtain essentially identical behavior for finite line lengths, ℓ_l, as long as ℓ_l is large compared to the mean value of the Wigner distribution, $\langle \Delta c\rangle$.) Also displayed in this figure (circles) is the distribution of the number of characters pushed into line 1000, c_{1000}. This clearly follows the same functional form. From our above considerations of the diffusion equation we see that $N(c_{1000}) \rightarrow f(x,t=1000)$, and that therefore the mean number of characters entered between catastrophic cascades is

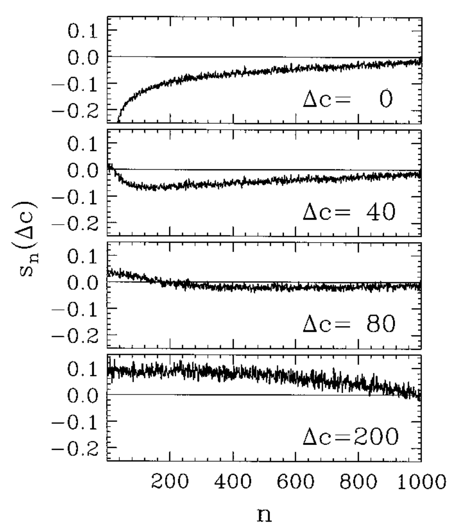

FIG. 4. Average stress, $s_n(\Delta c)$, as a function of the line number, n, and the number of characters entered since the last catastrophic cascade, Δc.

$$\langle \Delta c \rangle = \sqrt{\pi D n}. \tag{12}$$

We thus see that both distributions displayed in Fig. 2 are governed by diffusion physics.

Since Δc represents the number of characters entered between catastrophic cascades, and c_{1000} is the number of characters removed by a catastrophic cascade, sum rules require that $N(\Delta c)$ and $N(c_{1000})$ have the same norm and mean. The surprising aspect of Fig. 2 is that both distributions are identical, even down to the quadratic rise near the origin (see inset). The connection with the diffusion equation explains the Wigner-distribution form of $N(c_{1000})$, but not for $N(\Delta c)$.

Despite the same functional shape, the above two distributions are not tightly correlated. Figure 3 shows a density plot of the number distribution $N(c_{1000},\Delta c)$, where the gray level is proportional to the number of counts in a given bin. One sees only a weak enhancement of this number distribution along the diagonal. Here we plot the correlation between c_{1000} and the time delay to the next catastrophic cascade, but we obtain virtually identical results when plotting the correlation between c_{1000} and the time delay since the previous catastrophic cascade.

The total number of trailing blanks summed over all lines up to a certain maximum (here: 1000) changes by -1 each time a new character is entered. Catastrophic cascades change the total number of blanks by c_{1000}. Thus the total number of trailing blanks has exactly identical time dependence (up to a minus sign) as the total mass of the sandpile measured in Ref. [12].

In order to understand the trends of the buildup and release we compute the average stress in line n,

$$s_n(\Delta c) = \{\langle b_n \rangle - b_n\}_{\Delta c}, \tag{13}$$

as a function of the time delay, Δc, since the last catastrophe. (Here, the notation $\{\ \ \}_{\Delta c}$ indicates averaging over all events with identical value of Δc.) This is done in Fig. 4. We see from our data in Fig. 4 that a catastrophic cascade typically inserts more extra blanks in the early lines, thus reducing the stress in them below the average level. This indicates that large positive steps early in the random walk are correlated with catastrophic events. However, the structure of the stress as a function of the line number demonstrates the complex nature of the evolution of the self-organized critical state. This behavior cannot be explained in terms of a simple random walk.

In conclusion, we have examined self-organized criticality in line-wrap cascades in word processors. We find that the distribution of cascade lengths and cascade strengths can be accurately modeled with the diffusion equation and compared to analytic forms. Thus our system provides a connection between the exciting new field of complexity and somewhat more established branches of statistical physics. We find that the issue of predictability is complex. Although the distribution of times between large cascades is of a simple Wigner-distribution form, stress develops in a rather complicated style. Even though the present system represents a very simple nontrivial example of self-ordered criticality, the model inspires a wealth of questions, several of which we have addressed analytically and many more which remain unresolved, such as what is the optimum way to predict the onset of large cascades. By conquering this easily modeled example, insight may be reached regarding more complex systems such as sandpiles or more pertinent problems such as the modeling of earthquakes.

It is also possible to extend the model to higher dimensionality or to incorporate the addition of stress all through the paragraph rather than only through the first line. Work in these directions is currently in progress.

This research was supported by NSF Grant No. PHY-9403666, and W.B. was supported by NSF Grant No. PHY-9253505.

[1] P. Bak, C. Tang, and K. Wiesenfeld, Phys. Rev. Lett. **59**, 381 (1987).

[2] P. Bak and K. Chen, Sci. Am. **264**(1), 46 (1991).

[3] P. Bak and M. Creutz, in *Fractals in Science*, edited by A. Bunde and S. Havlin (Springer, Heidelberg, 1994), p. 26.

[4] M. Bretz, J.B. Cunningham, P.L. Kurczynnski, and F. Nori, Phys. Rev. Lett. **69**, 2431 (1992); J. Rosendahl, M. Vekii, and J. Kelley, Phys. Rev. E **47**, 1401 (1993); S.K. Grumbacher, K.M. McEwen, D.A. Halvorsen, D.T. Jacobs, and J. Lindler, Am. J. Phys. **61**, 329 (1993).

[5] B. Gutenberg and C. F. Richter, Ann. Geofis. **9**, 1 (1956).

[6] J. M. Carlson and J.S. Langer, Phys. Rev. Lett. **62**, 2632 (1989).

[7] D. Dhar and R. Ramaswamy, Phys. Rev. Lett. **63**, 1659 (1989); D. Dhar, *ibid.* **64**, 1613 (1990).

[8] T. Hwa and M. Kardar, Phys. Rev. A **45**, 7002 (1992); J.-P. Bouchaud, M.E. Cates, J. Rave Prakash, and S.F. Edwards, Phys. Rev. Lett. **74**, 1982 (1995).

[9] L.P. Kadanoff, S.R. Nagel, L. Wu, and S. Zhou, Phys. Rev. A **39**, 6524 (1989); J.M. Carlson, J.T. Chayes, E.R. Grannan, and G.H. Swindle, *ibid.* **42**, 2467 (1990); L.P. Kadanoff, A.N. Chhabra, A.J. Kolan, M.J. Feigenbaum, and I. Procaccia, *ibid.* **45**, 6095 (1992); K. Leung, *ibid.* **46**, 6211 (1992); V. Frette, Phys. Rev. Lett. **70**, 2762 (1993).

[10] P. Bak, K. Chen, and C. Tang, Phys. Lett. A **147**, 297 (1988); B. Drossel and F. Schwabl, Phys. Rev. Lett. **69**, 1629 (1992); V. Loreto, L. Pietronero, A. Vespignani, and Zapperi, *ibid.* **75**, 465 (1995).

[11] P. Bak and K. Sneppen, Phys. Rev. Lett. **71**, 4083 (1993); S. Maslov, M. Paczuski, and P. Bak, *ibid.* **73**, 2162 (1994).

[12] J. Rosendahl, M. Vekić, and J. E. Rutledge, Phys. Rev. Lett. **73**, 537 (1994).