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.
power-law limit can be shown to be and consequently the average number of trailing blanks in a line is

\[ \langle \Delta b \rangle = \frac{2\langle w \rangle - 2}{2\langle w \rangle - 3\langle w \rangle} \]

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 \(\langle w \rangle = 6\). A total of \(3 \times 10^6\) 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),

\[ N(n) \propto n^{-3/2}. \]  

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 \(\ell\) from line \(n\) to \(n+1\) increases the number of trailing blanks in line \(n\) by \(\ell\). Conversely, pushing a different word of length \(\ell'\) from line \(n-1\) into line \(n\) decreases the number of trailing blanks in line \(n\) by \(\ell'\). The probability distribution for a change \(\Delta b\) in the number of trailing blanks in a line is then

\[ P(\Delta b) = \sum_{b=0}^{\infty} p(b) p(b-\Delta b) \]  

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

\[ c_n = \sum_{i=1}^{n-1} \Delta b_i, \]  

where \(\Delta b_i\) is the change of the number of trailing blanks in line \(i\). If, for any \(n\), we have \(c_n < b_n\), where \(b_n\) is 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

\[ \partial_t f(x,t) = D \partial_x^2 f(x,t) \]  

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

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

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),

\[ D = \langle \Delta b_n^2 \rangle / 2. \]  

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

\[ J(t) = D \partial_x f(x,t) \big|_{x=0} = t^{-3/2}, \]  

in agreement with the numerical finding of Fig. 1.
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, \( l \), will result in a completely new first line, pushing the old first line into the second, and so on. Thus we get catastrophic cascades of cascades involving all lines—1000 in the specific example considered here—at least every \( l \) characters.

In Fig. 1(b) we show our results using identical parameters to Fig. 1(a), but using finite line lengths, \( l = 20, 30, 50, 70, 100, \) 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 \( \approx (\langle c \rangle_d)^{-2} \). 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 \( c_n \leq \langle c \rangle_d \) \( \forall n \). The corresponding solution to the diffusion equation is

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

with long-time behavior dominated by \( k_1 \), where \( k_j = j\pi \langle c \rangle_d \). For large word lengths, we have \( D \approx \langle c \rangle_d^2 \) and the exponent thus is \( \approx (\langle c \rangle_d\langle c \rangle_d)^{-2} \).

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, \( \Delta c \), 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 \( \Delta c \), \( N(\Delta c) \). It is obvious from this figure that they follow a Wigner distribution,

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

(As a side note we mention here that we obtain essentially identical behavior for finite line lengths, \( \langle c \rangle_d \), as long as \( \langle c \rangle_d \) 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

\[
 f_j(x,t) = \sum_{j=1}^{\infty} C_j \sin(k_j x) \exp(-Dk_j^2t) \tag{10}
\]
\[ \langle \Delta c \rangle = \sqrt{\pi D n}. \]  

(12)

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

Since \( \Delta c \) 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 \( n \),

\[ s_n(\Delta c) = \{ (b_n) - b_n \}_\Delta c, \]  

(13)
as a function of the time delay, \( \Delta c \), since the last catastrophe. (Here, the notation \( \{ \} \) indicates averaging over all events with identical value of \( \Delta c \).) 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.