Highly Optimized Tolerance: A Mechanism for Power Laws in Designed Systems

J. M. Carlson
Department of Physics, University of California, Santa Barbara, CA 93106

John Doyle
Control and Dynamical Systems, California Institute of Technology, Pasadena, CA 91125

(April 27, 1999)

We introduce a mechanism for generating power law distributions, referred to as *highly optimized tolerance* (HOT), which is motivated by biological organisms and advanced engineering technologies. Our focus is on systems which are optimized, either through natural selection or engineering design, to provide robust performance despite uncertain environments. We suggest that power laws in these systems are due to tradeoffs between yield, cost of resources, and tolerance to risks. These tradeoffs lead to highly optimized designs that allow for occasional large events. We investigate the mechanism in the context of percolation and sand pile models in order emphasize the sharp contrasts between HOT and self organized criticality (SOC), which has been widely suggested as the origin for power laws in complex systems. Like SOC, HOT produces power laws. However, compared to SOC, HOT states exist for densities which are higher than the critical density, and the power laws are not restricted to special values of the density. The characteristic features of HOT systems include: (1) high efficiency, performance, and robustness to designed-for uncertainties, (2) hypersensitivity to design flaws and unanticipated perturbations, (3) nongeneric, specialized, structured configurations, and (4) power laws. The first three of these are in contrast to the traditional hallmarks of criticality, and are obtained by simply adding the element of design to percolation and sand pile models, which completely changes their characteristics.

PACS numbers: 05.40.+j, 64.60.Ht, 64.60.Lx, 87.22.As, 89.20.+a

I. Introduction

One of the most pressing scientific and technological challenges we currently face is to develop a more complete and rigorous understanding of the behaviors that can be expected of complex, interconnected systems. While in many cases properties of individual components can be well characterized in a laboratory, these isolated measurements are typically of relatively little use in predicting the behavior of large scale interconnected systems or mitigating the cascading spread of damage due to the seemingly innocuous breakdown of individual parts. These failures are of particular concern due to the enormous economic, environmental, and/or social costs that often accompany them. This has motivated an increasing intellectual investment in problems which fall under the general heading of complex systems.

However, what a physicist refers to as a complex system is typically quite different from the complex systems which arise in engineering or biology. The complex systems studied in physics [1] are typically homogeneous in their underlying physical properties or involve an ensemble average over quenched disorder which is featureless on macroscopic scales. Complexity is associated with the emergence of dissipative structures in driven nonequilibrium system [2]. For a physicist complexity is most interesting when it is not put in by hand, but rather arises as a consequence of bifurcations or dynamical instabilities, which lead to emergent phenomena on large length scales.

This perspective is the driving force behind the concepts of self-organized criticality (SOC), introduced by Bak, Tang, and Wiesenfeld [3,4] and the edge of chaos (EOC) introduced by Kauffman [5] which have been the starting point for much of the interdisciplinary work on complex systems developed at the Santa Fe Institute and elsewhere. These theories begin with the idea that many complex systems naturally reside at a boundary between order and disorder, analogous to a bifurcation point separating a simple predictable state from fully developed chaos, or a critical point in equilibrium statistical physics. In these scenarios, there is a key state parameter, or density, which characterizes the otherwise generic, random, underlying system. In model systems, the density evolves self-consistently and without feedback to the specific value associated with the transition. Once at this point, large fluctuations inevitably emerge and recede as expected in the neighborhood of a second order transition. This gives rise to self-similarity, power laws, universality classes, and other familiar signatures of criticality. The widespread observations of power laws in geophysical, astrophysical, biological, engineered, and cultural systems has been widely promoted as evidence for SOC/EOC [6–13].

However, while power laws are pervasive in complex interconnected systems, criticality is not the only possible origin of power law distributions. Furthermore, there is little, if any, compelling evidence which supports other aspects of this picture. In engineering and biology, complex systems are almost always intrinsically *complicated,* and involve a great deal of built in or evolved structure and redundancy in order to make them behave in a reasonably predictable fashion in spite of uncertainties in their environment. Domain experts in areas such as biology and epidemiology, aeronautical and automotive design, forestry and environmental studies, the Internet, traffic, and power systems, tend to reject the concept of universality, and instead favor descriptions in which the detailed structure and external conditions are key factors in determining the performance and reliability of their systems. The complexity in designed systems often leads to apparently simple, predictable, robust behavior. As a result, designed complexity becomes increasingly hidden, so that its role in determining the sensitivities of the system tends to be underestimated by nonexperts, even those scientifically trained.

The Internet is one example of a system which may superficially appear to be a candidate for the self-organizing theory of complexity, as power laws are ubiquitous in Internet statistics [14,15]. It certainly appears as though new users, applications, workstations, PCs, servers, routers, and whole subnetworks can be added and the entire system naturally self-organizes into a new, robust configuration. Furthermore, once on-line, users act as individual agents, sending and receiving messages according to their needs. There is no centralized control, and individual computers both adapt their transmission rates to the current level of congestion, and recover from network failures, all without user intervention or even awareness. It is thus tempting to imagine that Internet traffic patterns can be viewed as an emergent phenomena from a collection of independent agents who adaptively self-organize into a complex state, balanced on the edge between order and chaos, with ubiquitous power laws as the classic hallmarks of criticality.

As appealing as this picture is, it has almost nothing to do with real networks. The reality is that modern internets use sophisticated multi-layer protocols [16] to create the illusion of a robust and self-organizing network, despite substantial uncertainty in the user-created environment as well as the network itself. It is no accident that the Internet has such remarkable robustness properties, as the Internet protocol suite (TCP/IP) in current use was the result of decades of research into building a nationwide computer network that could survive deliberate attack. The high throughput and expandability of internets depend on these highly structured protocols, as well as the specialized hardware (servers, routers, caches, and hierarchical physical links) on which they are implemented. Yet it is an important design objective that this complexity be hidden.

The core of the Internet, the Internet Protocol (IP), presents a carefully crafted illusion of a simple (but possibly unreliable) datagram delivery service to the layer above (typically the Transmission Control Protocol, or TCP) by hiding an enormous amount of heterogeneity behind a simple, very well engineered abstraction. TCP in turn creates a carefully crafted illusion to the applications and users of a reliable and homogeneous network. The internal details are highly structured and non-generic, creating apparent simplicity, exactly the opposite from SOC/EOC. Furthermore, many power law statistics of the Internet are independent of density (congestion level), which can vary enormously, suggesting that criticality may not be relevant.

Interestingly and importantly, the increase in robustness, productivity, and throughput created by the enormous internal complexity of the Internet and other complex systems is accompanied by new hypersensitivities to perturbations the system was not designed to handle. Thus while the network is robust to even large variations in traffic, or loss of routers and lines, it has become extremely sensitive to bugs in network software, underscoring the importance of software reliability and justifying the attention given to it. The infamous Y2K bug, though not necessarily a direct consequence of network connectivity, is nevertheless the best-known example of the general risks of high connectivity for high performance. There are many more less well-known examples, and indeed most modern large-scale network crashes can be traced to software problems, as can the failures of many systems and projects (eg. the Ariane 5 crash or the Denver Airport Baggage handling system fiasco). We will return to the Internet and other examples at the end of the paper.

This "robust-yet-fragile" feature is characteristic of complex systems throughout engineering and biology. If we accept the fact that most real complex systems are highly structured, dominated by design, and sensitive to details, it is fair to ask whether there can be any meaningful theory of complex systems. In other words, are there common features, other than power laws, that the complicated systems in engineering and biology share that we might hope to capture using simple models and general principles? If so, what role can physics play in the development of the theory?

In this paper we introduce an alternative mechanism for complexity and power laws in designed systems which captures some of the fundamental contrasts between designed and random systems mentioned above in simple settings. Our mechanism leads to (1) high yields robust to designed-for uncertainty, (2) hypersensitivity to design flaws and unanticipated perturbations, (3) stylized and structured configurations, and (4) power law distributions. These features arise as a consequence of optimizing a design objective in the presence of uncertainty and specified constraints. Unlike SOC or EOC, where the external forces serve only to initiate events and the mechanism which gives rise to complexity is essentially self-contained, our mechanism takes into account the fact that designs are developed and biological systems evolve in a manner which rewards successful strategies subject to a specific form of external stimulus. In our case uncertainty plays the pivotal role in generating a broad distribution of outcomes. We somewhat whimsically refer to our mechanism as *highly optimized tolerance* (HOT), a terminology intended to describe systems which are designed for high performance in an uncertain environment.

The specific models we introduce are not intended as realistic representations of designed systems. Indeed, in specific domain applications at each level of increased model sophistication, we expect to encounter new structure which is crucial to the robustness and predictability of the system. Our goal is to take the first step towards more complicated structure in the context of familiar models to illustrate how even a small amount of design leads to significant changes in the nature of an interconnected system. We hope that our basic results will open up new directions for the study of complexity and cascading failure in biological and engineering systems.

To describe our models, we will often use terminology associated with a highly simplified model of a managed forest which is designed to maximize timber yield in the presence of fire risk. Suppose that in order to attain this goal, the forester constructs firebreaks at a certain cost per unit length, surrounding regions that are expected to be most vulnerable (e.g., near roads and populated areas or tops of hills where lightning strikes are likely). At best, this is remotely connected to real strategies used in forestry [17,18]. Our motivation for using a "forest fire" example is the familiarity of similar toy models in the study of phase transitions and SOC [19].

The optimal designed toy forest contains a highly stylized pattern of firebreaks separating high density forested regions. The regions enclosed by breaks are tailored to the external environment and do not resemble the fractal percolation-like clusters of the forest fire model which has been studied in the context of SOC. Furthermore, there is nothing in the designed forest resembling a critical point. Nonetheless, the relationship between the frequency and size of fires in designed systems is typically described by a power law. In an optimized design, firebreaks are concentrated in the regions which are expected to be most vulnerable, leaving open the possibility of large events in less probable zones.

The forest fire example illustrates the basic ingredients of the mechanism for generating power laws which we describe in more detail below. If the trees were randomly situated with a comparable density to that of the designed system, any fire once initiated would almost surely spread throughout the forest generating a systemwide event. Designed configurations represent very special choices and comprise a set of measure zero within the space of all possible arrangements at a given density. Systems are tuned to highly structured and efficient operating states either by deliberate design or evolution by natural selection. In contrast, in SOC large connected regions emerge and recede in the dynamically evolving statistically steady state where no feedback is incorporated to set the relative weights of different configurations.

In the sections that follow, we use a variety of different model systems and optimization schemes to illustrate properties of the HOT state. These include a general argument for power laws in optimized systems based on variational methods (Section II), as well as numerical and analytical studies of lattice models (Sections III-VI). In an effort to clarify the distinctions between HOT and criticality (summarized in Section V), we introduce variants of familiar models from statistical physics (Section III)– percolation with sparks and the original sand pile model introduced by Bak, Tang, and Wiesenfeld [3]. Both models are modified to incorporate elementary design concepts, and are optimized for yield Y in the presence of constraints. In percolation, yield is the number of occupied sites which remain after a spark hits the lattice and burns all sites in the associated connected cluster. In the designed sand piles, yield is defined to be the sand untouched by an avalanche after a single grain is added to the system. When we introduce design, these two problems become essentially identical, and optimizing yield leads us to construct barriers which minimize the expected size of the event based on a prescribed density for the spatial dependence of the probability of triggering events. In this way we mimic engineering and evolutionary processes which favor designs that maximize yield in the presence of an uncertain environment. We consider both a global optimization over a constrained subclass of configurations (Section IV), as well as a local, incremental algorithm which develops barriers through evolution (Section VI). We conclude with a summary of our results, and a discussion of a few specific applications where we believe these ideas may apply.

II: Power Laws and Design

If the power laws in designed systems arise due to mechanisms entirely unlike those in critical phenomena then the ubiquity of power laws needs a fresh look. If engineering systems could be constructed in a self-similar manner it would certainly simplify the design process. However, self-similar structures seldom satisfy sophisticated design objectives. With the exception of distribution networks which are inherently tree-like and often fractal, hierarchies of subsystems in complex biological and engineering systems have a self-dissimilar structure. For example, organisms, organs, cells, organelles, and macromolecules all have entirely different structure [20]. The hundreds of thousands of subsystems in a modern commercial aircraft do not themselves resemble the full aircraft in form or function, nor do their subsystems, and so on. Thus if power laws arise in biological and engineering systems, we would not necessarily expect that they would be connected with self-similar structures, and our idealized designed systems in fact turn out to be self-dissimilar.

We begin our analysis with a general argument for the presence of heavy tails in the distribution of events which applies to a broad class of designed systems. Consider an abstract $d-$dimensional space denoted by X which acts as a substrate for events in our system. This can be thought of concretely as a forest, where the coordinates of the trees, firebreaks, and sparks which initiate fires are defined in X. Alternately, X could correspond to an abstract map of interconnected events in which a failure at one node may trigger failures at connected nodes. We assume there is some knowledge of the spatial distribution of probabilities of initiating events (sparks), and some resource (firebreaks) which can be used to limit the size of events (fires). There is some cost or constraint associated with use of the resource, and an economic gain (i.e. increased yield) associated with limiting the sizes of events.

We define $p(\mathbf{x})$ to be the probability distribution for initiating events $\forall \mathbf{x} \in X$. Let $A(\mathbf{x})$ denote the size of the region which experiences the event initiated at $\mathbf{x}$, and let cost $C(x)$ scale as $A^\alpha(\mathbf{x})$. In general α will be a positive number which sets the relative weight of events of different sizes. If we are simply interested in the area of the region then $\alpha = 1$. For cases in which X is continuous, the expected cost of the avalanche is given by:

$$E(A^\alpha) = \int_X p(\mathbf{x}) A^\alpha(\mathbf{x}) d\mathbf{x}. \tag{1}$$

Let $R(\mathbf{x})$ denote the resource which restricts the sizes of the events. Constraints on $R(\mathbf{x})$ can take a variety of forms. Here we consider the simplest case which corresponds to a limitation on the total quantity of the resource,

$$\int_X R(\mathbf{x}) d\mathbf{x} = \kappa \tag{2}$$

where κ is a constant. Alternatively, the constraint on $R(\mathbf{x})$ could be posed in terms of a fixed total number of regions within X, or a cost benefit function Q could be introduced balancing the benefit of a small expected size (Eq. (1)) with the cost associated with use of the resource.

We will assume that the local event size is inversely related to the local density or cost of the resource, so that $A(\mathbf{x}) = R^{-\beta}(\mathbf{x})$, where typically β is positive. This relationship arises naturally in systems with spatial geometry (e.g. in the forest fire analogy), where in d dimensions we can think of $R(\mathbf{x})$ as being $d-1$ dimensional separating barriers. In that case $A(\mathbf{x}) \sim R^{-d}(\mathbf{x})$. In some systems the relationship between $A(\mathbf{x})$ and $R(\mathbf{x})$ is difficult to define uniquely, and in some cases reduces to a value judgement. Here our spatially motivated assumption that $A(\mathbf{x}) = R^{-\beta}(\mathbf{x})$ is important for obtaining power law distributions. If we assume an exponential relationship between the size of an event and its cost (e.g. $A \sim -log(R)$), we obtain a sharp cutoff in the distribution of events. In essence, this is because it becomes extremely inexpensive to restrict large events because the cost of resources decreases faster than the size of the event to any power. Alternately, one could define a cost function for cases in which there is a large social or ethical premium (e.g. loss of life) associated with large events. This could lead to a cutoff in the distribution due to a rapid rise in the total allocation of resources to prevent large events. In this case, the heavy tails would occur in the cost C and not in the event size A.

To obtain the HOT state we simply minimize the expected cost (Eq. (1)) subject to the constraint (Eq. (2)). Substituting the relationship $A(\mathbf{x}) = R^{-\beta}(\mathbf{x})$ into Eq. (1) we obtain

$$E(A^\alpha) = \int_X p(\mathbf{x}) R^{-\alpha\beta}(\mathbf{x}) d\mathbf{x}. \tag{3}$$

Combining this with Eq. (2), we minimize $E(A^\alpha)$ using the variational principle by solving

$$\delta \int_X \left[p(\mathbf{x}) R^{-\alpha\beta}(\mathbf{x}) - \lambda R(\mathbf{x}) \right] d\mathbf{x} = 0. \tag{4}$$

Thus the optimal relationship between the local probability and constrained resource is given by

$$p(\mathbf{x}) R^{-\alpha\beta-1}(\mathbf{x}) = \text{constant}. \tag{5}$$

From this we obtain

$$p(\mathbf{x}) \sim R^{\alpha\beta+1}(\mathbf{x}) \sim A^{-(\alpha+1/\beta)}(\mathbf{x}) \sim A^{-\gamma}(\mathbf{x}), \tag{6}$$

where $\gamma = \alpha + 1/\beta$. This relation should be viewed as the local rule which sets the best placements of the resource. As expected, greater resources are devoted to regions of high probability.

As function of $\mathbf{x}$, Eq. (6) shows that $p(\mathbf{x})$ and $A(\mathbf{x})$ scale as a power law. However, we want to obtain the distribution $P(A)$ as a function of the area A rather than the local coordinate $\mathbf{x}$. It is convenient to focus on cumulative distribution, $P_{\text{cum}}(A)$, which is the sum of $P(A)$ for regions of size greater than or equal to A. We express the tails of $P_{\text{cum}}(A)$ as

$$\begin{aligned} P_{\text{cum}}(A) &= \textstyle\int_{A(\mathbf{x})>A} p(\mathbf{x}) d\mathbf{x} \\ &= \textstyle\int_{p(\mathbf{x})<A^{-\gamma}} p(\mathbf{x}) d\mathbf{x} \end{aligned} \tag{7}$$

where the integral is evaluated over the subset of $\mathbf{x}$ in which the local value $A(\mathbf{x})$ is greater than the specified value A.

Under what conditions does this relationship lead to heavy tails? Certainly not all $p(\mathbf{x})$ lead to power laws in $P(A)$ (equivalently, $P_{\text{cum}}(A)$, which has power law tails if $P(A)$ has power law tails, with one power higher in the exponent). For example, if $p(\mathbf{x})$ is concentrated within a finite region then the resource would optimally be concentrated within that region, and the distribution $P(A)$ would *a priori* have zero weight for events greater than the area associated with the mass concentration of $p(\mathbf{x})$. Here the most extreme case is a point mass at a particular location: $p(\mathbf{x}) = \delta(\mathbf{x} - \mathbf{x}^*)$, which could be enclosed by a high density of the resource, so that all activity is confined to $\mathbf{x}^*$. Alternately, if $p(\mathbf{x})$ is spatially uniform, then $R(\mathbf{x})$ and $A(\mathbf{x})$ would be uniformly distributed, and $P(A)$ would be a point mass at a fixed area determined by the resource constraint and the system size.

While counterexamples such as those we have just described can be constructed, a broad class of distributions $p(\mathbf{x})$ leads to heavy tails in $P(A)$. The case for $d = 1$ with monotonic $p(\mathbf{x})$ and restricting X to $\mathbf{x} > 0$ is particularly simple (and forms the basis for the more general case). In this special case, the change of variables from $p(\mathbf{x})$ to $P(A)$ is straightforward and we obtain

$$\begin{aligned} P_{\text{cum}}(A) &= \textstyle\int_{p^{-1}(A^{-\gamma})}^{\infty} p(\mathbf{x}) d\mathbf{x} \\ &= p_{\text{cum}}(p^{-1}(A^{-\gamma})) \end{aligned} \tag{8}$$

where $p_{\text{cum}}(\mathbf{x})$ is the tail of the cumulative distribution for the probability of hits and p^{-1} is the inverse function of p, so that $p^{-1}(A^{-\gamma})$ is the value of $\mathbf{x}$ for which $p(\mathbf{x}) = A^{-\gamma}$.

We can use Eq. (8) to directly compute the tail of $P_{\text{cum}}(A)$ for standard $p(x)$, such as power laws, exponentials, and Gaussians. Table 1 summarizes the results, where we look only at tails in the distributions of x and A, and drop constants. We get a power distribution for $P_{\text{cum}}(A)$ in each case, with a logarithmic correction for the Gaussian.

$p(x)$	$p_{\rm cum}(x)$	$P_{\rm cum}(A)$
$x^{-(q+1)}$	x^{-q}	$A^{-\gamma(1-1/q)}$
e^{-x}	e^{-x}	$A^{-\gamma}$
e^{-x^2}	$x^{-1}e^{-x^2}$	$A^{-\gamma}[\log(A)]^{-1/2}$

TABLE I. In the HOT state power law distributions of the region sizes $P_{\rm cum}(A)$ are obtained for a broad class of probability distributions of the hits $p(\mathbf{x})$, including power law, exponential, and Gaussian distributions as shown here.

For higher dimensions, suppose that the tails of $p(\mathbf{x})$ can be bounded above and below by

$$p_l(|\mathbf{x}|) \leq p(\mathbf{x}) \leq p_u(|\mathbf{x}|), \tag{9}$$

where $|\mathbf{x}|$ denotes the magnitude of $\mathbf{x}$. The specific form of Eq. (9) effectively reduces the change of variables to quasi-one-dimensional computations. With this assumption, Eq. (7) can be bounded below by

$$\begin{aligned} P_{\rm cum}(A) &\geq D \int_{p_l(\mathbf{x})<A^{-\gamma}} p_l(r) r^{d-1} dr & (10)\\ &\geq D \int_{p_l(\mathbf{x})<A^{-\gamma}} p_l(r) dr & (11) \end{aligned}$$

where D is a constant and the last inequality holds in the tails where $r > 1$. Provided p_u and p_l are asymptotically of the same order, this implies that the case of $d > 1$ simply adds additional weight to the tail. More detailed computations can be made to compute exactly what the $d > 1$ correction terms are for various distributions.

While this analysis is fairly abstract, the underlying concepts are highly intuitive, and the basic results should carry over to a wide variety of spaces, resources, and constraints. In essence we contend that optimizing yield will cause the design to concentrate protective resources where the risk of failures are high, and to allow for the possibility of large rare events elsewhere.

III: Lattice Models

In this section we consider two familiar lattice models from statistical physics, first as traditionally defined and then incorporating design. These include percolation [21], the simplest model which exhibits a second order phase transition, and the original sand pile model introduced by Bak, Tang, and Wiesenfeld [3]. In the context of optimization and design these two models become essentially identical, so we consider them together.

III.A Percolation

We begin with site percolation on a two dimensional $N \times N$ square lattice. In the random case, sites are occupied with probability p and vacant with probability $1-p$. For a given density $\rho = p$ all configurations are equally likely. Typical configurations have a random, unstructured appearance, as illustrated in Fig. 1a. At low densities, nearest neighbor occupied sites form isolated clusters. The distribution of cluster sizes cuts off sharply at a characteristic size which depends on density. The critical density p_c marks the divergence of the characteristic cluster size, and at p_c the cluster size distribution is given by a power law. Above p_c there is an infinite cluster which corresponds to a finite fraction of the system. At p_c the infinite cluster exists but is sparse, with a nontrivial fractal dimension. The percolation order parameter, $P_\infty(p)$ is the probability that any particular site is connected to the infinite cluster. For $p < p_c$, $P_\infty(p) = 0$. At $p = p_c$, $P_\infty(p)$ begins to increase monotonically from zero to unity at $p = 1$. In the neighborhood of the transition, the critical exponent β describes the onset of percolation: $P_\infty(p) \sim (p - p_c)^\beta$. An extensive discussion of percolation can be found in [21].

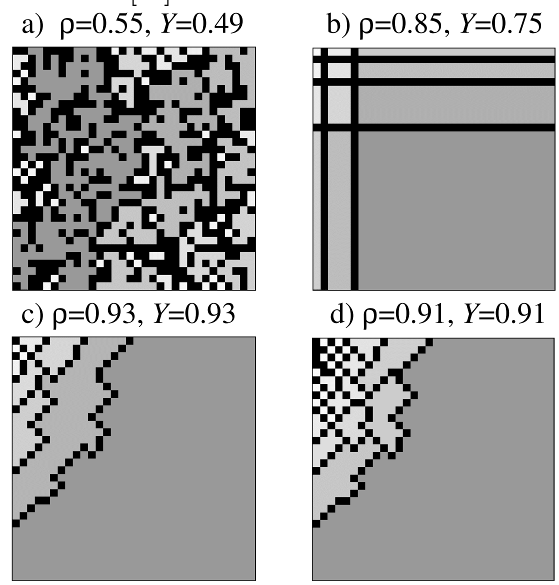

FIG. 1. Sample percolation configurations on a 32×32 lattice for (a) the random case near p_c, (b) a HOT grid (Section IV), and HOT states obtained by evolution (Section V) at (c) optimal yield, and (d) a somewhat lower density. Unoccupied sites are black, and clusters are grey, where darker shades indicate larger clusters. The designed systems are generated for an asymmetric distribution of hitting probabilities with Gaussian tails, peaked at the upper left corner of the lattice.

In order to introduce risk and compute yield, we define a very primitive dynamics in which for a given assignment of vacant and occupied sites, a single spark is dropped on the lattice initiating a fire. In the standard forest analogy, occupied sites correspond to trees, and risk is associated with fires. The yield Y is defined to be the average density of trees left unburnt after the spark hits. If a spark hits an unoccupied site, nothing burns. When the spark hits an occupied site the fire spreads throughout the associated cluster, defined to be the connected set of A nearest neighbor occupied sites.

We let $P(A)$ denote the distribution of events of size A, and let $P_{\rm cum}(A)$ denote the cumulative distribution of events greater than or equal to A. The yield is then $Y(\rho) = \rho - < P >$ where the average $< P >$ is computed with respect to both the ensembles of configurations and the spatial distribution $p(i,j)$ of sparks. By translation invariance, results for the random case are independent of the distribution of sparks.

In Fig. 2a we plot yield Y as function of the initial density ρ for a variety of different scenarios including both random percolation and design. The maximum possible yield corresponds to the diagonal line: $Y = \rho$, which is obtained if a vanishing fraction of the sites are burned after the spark lands. The yield curve for the random case is depicted by the dashed line in Fig. 2a. At low densities the results coincide with the maximum yield. Near $\rho = p_c$ there is a crossover, and $Y(\rho)$ begins to decrease monotonically with ρ, approaching zero at high density.

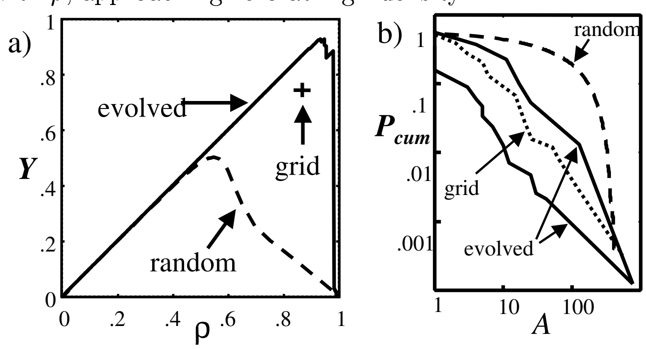

FIG. 2. Comparison between HOT states and random systems at criticality for the percolation model: (a) Yield vs. Density: $Y(\rho)$, and (b) cumulative distributions of events $P_{\rm cum}(A)$ for cases (a)-(d) in Fig. 1.

The crossover becomes sharp as $N \to \infty$ and is an immediate consequence of the percolation transition. In the thermodynamic limit only events involving the infinite cluster result in a macroscopic event. Yield is computed as the sum of contributions associated with cases in which (i) the spark misses the infinite cluster and the full density is retained, and (ii) the spark hits the infinite cluster, so that compared with the starting density the yield is reduced by the fraction associated with the infinite cluster:

$$Y(\rho) = [1 - P_\infty(\rho)]\rho + P_\infty(\rho)[\rho - P_\infty(\rho)]$$

$$= \rho - {P_\infty}^2(p). \qquad (12)$$

Thus yield is simply related to the percolation order parameter, and the exponent which describes the departure of yield from the maximum yield curve in the neighborhood of the transition is 2β. In random percolation, where the only tunable parameter is the density, the optimal yield coincides with the critical point.

III.B Sand Piles

Now we turn to the sand pile model, which was introduced by Bak, Tang, and Wiesenfeld (BTW) as the prototypical example of SOC. Unlike percolation, the sand pile model is explicitly dynamical. It is an open driven system which evolves to the critical density upon repeated iteration of the local rules.

The model is defined on an $N \times N$ integer lattice. The number of grains of "sand" on each site is given by $h(i,j)$. The algorithm which defines the model consists of the individual addition of grains to randomly selected sites:

$$h(i,j) \to h(i,j) + 1, \qquad (13)$$

such that the site (i,j) topples if the height exceeds a prescribed threshold h_c. As a result $h(i,j)$ is reduced by a fixed amount which is subsequently redistributed among nearest neighbor sites h_{nn}. We take $h_c = 4$ and the toppling rule

$$\begin{aligned} h(i,j) \geq h_c: \quad & h(i,j) \to h(i,j) - 4 \\ & h_{nn} \to h_{nn} + 1. \end{aligned} \qquad (14)$$

Sand leaves the system when a toppling site is adjacent to the boundary. The toppling rule is iterated until all sites are below threshold, at which point the next grain is added.

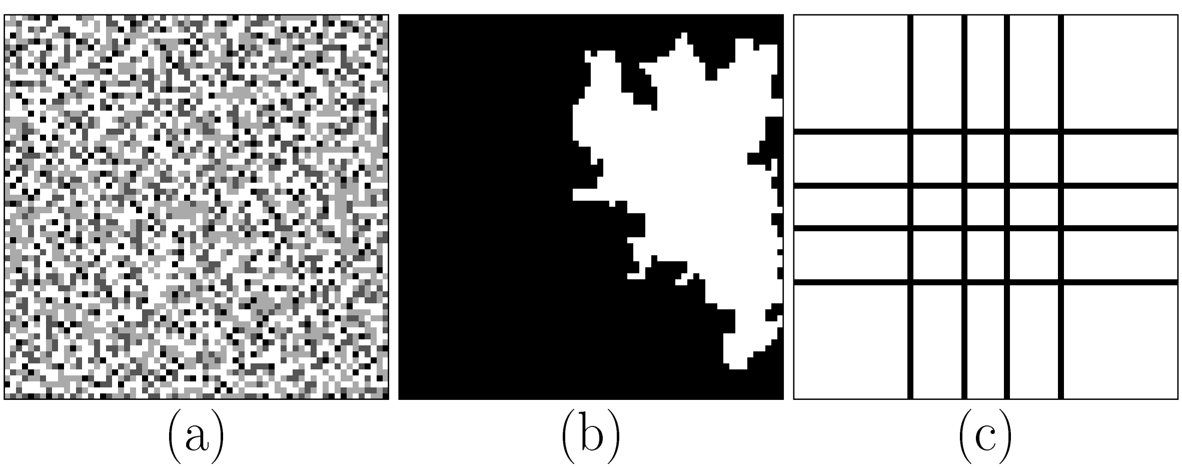

FIG. 3. SOC vs. HOT states in the BTW sand pile model with $N = 64$. The grayscale ranges from black ($h = 0$) to white ($h = 3$). Figure (a) is typical SOC configuration, and (b) illustrates (in white) the area swept out by a typical (fractal) large event. Figure (c) illustrates the HOT state for a grid design with 4 horizontal and vertical cuts, and a symmetric Gaussian distribution ($\sigma = 10$) of hits. Here the area swept out by events are the rectangular regions delineated by cuts.

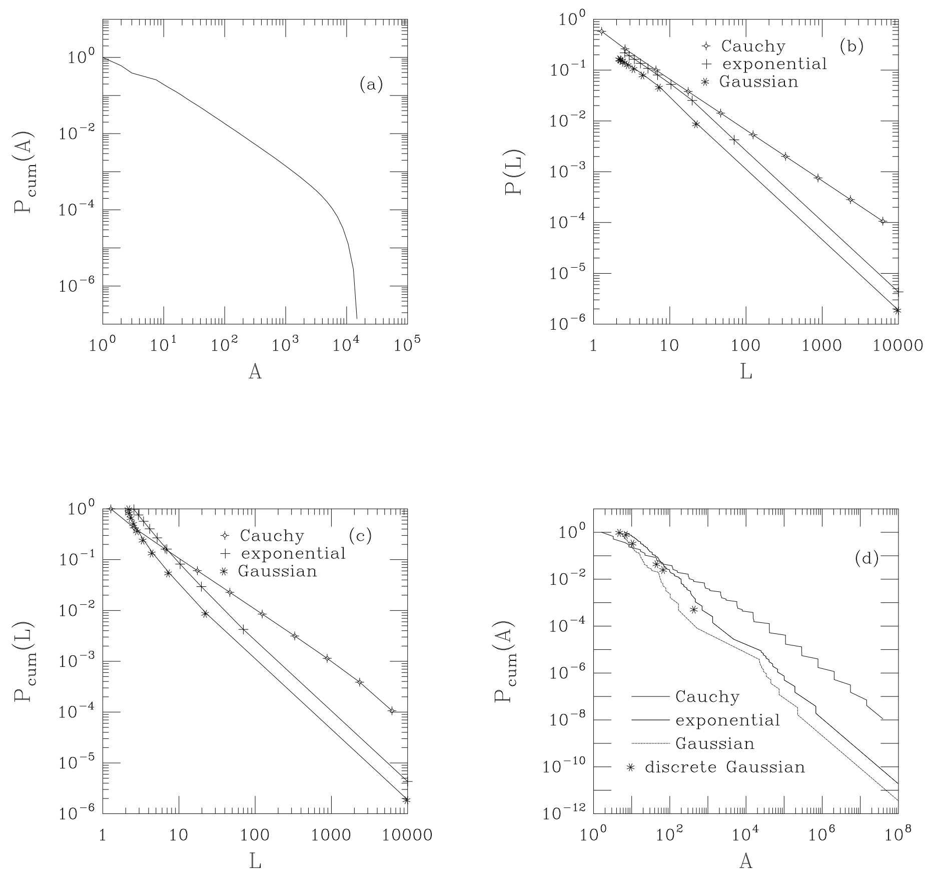

FIG. 4. Both SOC and HOT states exhibit power laws in the avalanche distributions. In (a), (c), and (d) we plot the distributions for the probability $P_{\text{cum}}(A)$ of observing an event of size greater than or equal to A. Figure (a) illustrates results for the 128×128 BTW sand pile. Figure (b-d) illustrates results for the HOT state in the continuum limit. Results are obtained for Cauchy, exponential and Gaussian distributions of hits (see text). Figure (b) illustrates $P(L)$ vs. L for $d = 1$. Figure (c) shows the corresponding cumulative distributions. Figure (d) shows the cumulative distribution of areas for $d = 2$, obtained by overlaying the $d = 1$ solutions. Numerical results for a 512×512 discrete lattice with 4 horizontal and 4 vertical cuts are included for comparison with the Gaussian case.

Despite the apparent simplicity of the algorithm, this and related SOC models exhibit rich dynamics. The BTW model does not exhibit long range height correlations [25] (Fig. 3a illustrates a typical height configuration), but it still exhibits power laws in the distribution of sizes of the avalanches. Here size is defined to be the number of sites which topple as the result of the addition of a single grain to the pile (see Fig. 4a). In addition, the model exhibits self-similarity in certain spatial and temporal features such as fractal shapes of the individual regions which exhibit avalanches (see Fig. 3b) and power law power spectra of the time series of events.

Like equilibrium systems, such as the random percolation model in the neighborhood of a critical point, SOC systems exhibit no intrinsic scale. The power law describing the distribution of sizes of events extends from the microscopic scale of individual sites out to the system size (see Fig. 4a). Indeed, for some SOC models concrete mappings to equilibrium critical points have been obtained [22–24]. In the BTW sand pile model, the critical point is associated with a critical density (average height) of sand on the pile of roughly $< h >_c = 2.125$.

We define yield for the sand pile model to be the number of grains left untouched by an avalanche following the addition of a single grain. That is, once the system has reached a statistically steady state, we compute yield for a given configuration after one complete iteration of the addition (Eq. (13)) and toppling (Eq. (14)) rules, as the sum of heights over the subset of sites U which are not hit during that particular event, and then average the result over time:

$$Y(\rho) = < N^{-2} \sum_U h(i,j) > \tag{15}$$

The result is illustrated in Fig. 5. For the SOC system computing yield as a time average of iterative dynamics is equivalent to computing an ensemble average over different realizations of the randomness. The results are insensitive to changes in the spatial distribution of addition sites. Essentially the same event size distributions are obtained regardless of whether grains are added at a particular site, a subset of sites, or randomly throughout the system.

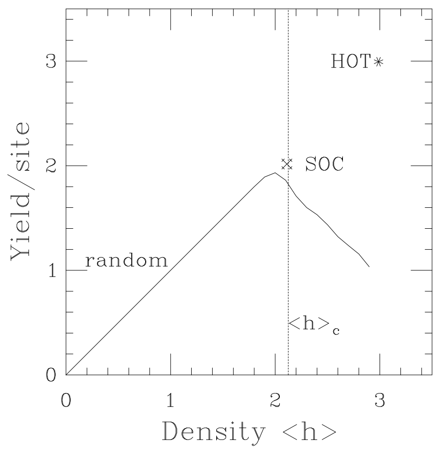

FIG. 5. Yield vs. density. We compare the yield (defined to be the number of grains left on those sites of the system which were unaffected by the avalanche) for different ways of preparing the system. Results are shown for randomly generated stable initial conditions, which are subject to a single addition (solid line) for a 128×128 sand pile model, and the corresponding SOC state and the HOT state. Clearly the HOT state outperforms the other systems, exhibiting a greater yield at higher density. Yield in the HOT state can be made arbitrarily close to the maximum value of 3 for large systems with a sufficient number of cuts, while increasing system size does not significantly alter the yield in the other two cases.

Unlike random percolation, in which we obtained a one parameter curve describing yield as a function of density, our result for the sand pile model corresponds to a single point because the mean density $< h_c >$ reaches a steady state. However, it is possible to make a more direct connection between our results for the sand pile model and percolation, by considering a modified sand pile model in which the density is an adjustable parameter. Aside from a few technical details, this coincides with the closed, equilibrium analog of the sand pile model mentioned above. Alternately, it can be thought of as a primitive, one parameter, probabilistic design.

Suppose we can manipulate a single degree of freedom, the density of the initial state. That is, we begin with an empty lattice, and add grains randomly until the systemwide density achieves the value we prescribe. We also restrict all initial heights to be below threshold. This results in a truncated binomial distribution of heights, restricted to values $h(i,j) \in [0,1,2,3]$, where the mean is adjusted to produce the prescribed density. In Fig. 5 we compute the mean yield vs. density of this system after one grain is added, as an average over both the initial states and the random perturbation sites. As in percolation, densities near the critical point produce the maximum yield. Systems which are designed at low densities are poor performers in terms of the number of grains left on the system after an avalanche because so few grains were there in the first place. At the critical density, the characteristic size of the avalanche triggered by the perturbation becomes of order the system size. Densities beyond the critical density often lead to systemwide events, causing the yield to drop. In fact, both the peak density and yield of the primitive design are nearly equal to the time averaged yield and density of the SOC state [25], where for each event the yield is the total number of grains left on sites which do not topple.

It is important to note that the primitive design is not equivalent to SOC. The mechanisms which lead the system to the critical density are entirely different in the two cases. In SOC the critical density is the global attractor of the dynamics, which follows from the fact that the system is driven at an infinitesimal rate. In contrast, the primitive design is tuned (by varying the density) to obtain maximum yield. Consequently, the primitive design has statistics which mimic SOC in detail, but without any "self-organization." Thus it would be difficult to distinguish on the basis of statistics alone whether a system exhibits SOC or is merely a manifestation of a primitive design process.

III.C HOT States

In this subsection we show that it is possible to retain maximum yields well beyond the critical point, and up to the maximum density as $N \to \infty$. This is made possible by selecting a measure zero subset of *tolerant* states. We refer to these sophisticated designs as HOT states, because we fix the exact configuration of the system, laying out a high density pattern which is robust to sparks or the addition of grains of sand.

In our designed configurations, in most respects there will be no distinction between a designed percolation configuration and a designed sand pile. In percolation, densities well above the critical density are achieved by selecting configurations in which clusters of occupied sites are compact. In the sand pile model we construct analogous compact regions in which most sites are chosen to be one notch below threshold: $h(i,j) = h_c - 1 = 3$, which are analogous to the occupied sites in percolation. In each case to limit the size of the avalanches, barriers of unoccupied sites, or sites with $h(i,j) = 0$ are constructed, which, as discussed in Section II, are subject to a constraint.

As stated previously in Section II, the key ingredients for identifying HOT states are the probability distribution of perturbations, or sparks, $p(i,j)$, and a specification of constraints on the optimization or construction of barriers. We will begin by considering a global optimization over a restricted subclass of configurations. Numerical and analytical results for this case are obtained in Section IV. In Section V we introduce a local incremental optimization scheme, which is reminiscent of evolution by natural selection. Sample HOT states are illustrated in Figs. 1 and 3.

In the grid design, we define our constraint such that the boundaries are composed of horizontal and vertical cuts. For percolation, the cuts correspond to lines comprised of unoccupied sites. In the sand pile model the cuts correspond to lines along which $h(i,j)=0$. In the sand pile model, somewhat higher yields are obtained if the cuts are defined to have height 2, and contiguous barriers of height two are also sufficient to terminate an avalanche when the BTW toppling rule is iterated. However, the difference in density between a grid formed with cuts of height zero and 2 is a finite size effect which does not alter the event size distribution, and leads to a system which is less robust to multiple hits.

A set of $2(n-1)$ cuts $\{i_1, i_2, ...i_{n-1}, j_1, j_2, ...j_{n-1}\}$ defines a grid of n^2 regions on the lattice. For a given configuration (set of cuts), the distribution of event sizes and ultimately the yield are obtained as an ensemble average. The system is always initialized in the designed state. Event sizes are determined by the enclosed area and contribute to the distribution with a weight determined by the sum of the enclosed probability $p(i,j)$.

IV: Optimization of the Grid Design

For the grid configurations (Figs. 1b and 3c), the design problem involves choosing the optimal set of cuts which minimizes the expected size of the avalanche. First we consider two simple cases. Suppose you know exactly which site (i,j) will receive the next grain. Then clearly the best strategy is to define one of the cuts to coincide with that site, so that when a grain is added to the system the site remains sub-threshold and no avalanche occurs. Alternatively, if $p(i,j)$ is spatially uniform, then the best design strategy is to define equally spaced cuts: $i_1 = N/n, i_2 = 2N/n, ..., i_{n-1} = (n-1)N/n, j_1 = N/n, ...j_{n-1} = (n-1)N/n$, so that the system is divided into n^2 regions of equal area. In this case, all avalanches are of size $(N/n)^2$. Already we see that the avalanche size is considerably less than that which would be obtained in the SOC or percolation models at the same density (the SOC system will never attain the high densities of the HOT state).

The more interesting case arises when you have some knowledge of the spatial distribution of hitting probabilities. For a specified set of cuts the expected size of the avalanche (defined to be the number of toppling sites) is given by

$$E(A) = \sum_{\mathcal{R}} \mathcal{P}(\mathcal{R}) A(\mathcal{R}), \tag{16}$$

where for a given set of horizontal and vertical cuts the sum is over the rectangular regions $\mathcal{R}$ of the grid, and $\mathcal{P}(\mathcal{R})$ and $A(\mathcal{R})$ represent the cumulative probability and total area of region $\mathcal{R}$ defined generally on a d-dimensional space X as

$$\mathcal{P}(\mathcal{R}) = \int_{\mathcal{R}} p(\mathbf{x}) d\mathbf{x} \quad \text{and} \quad A(\mathcal{R}) = \int_{\mathcal{R}} d\mathbf{x}. \tag{17}$$

Equation (16) can be written in terms of the hitting probability distribution $p(i,j)$ and the positions of the i and j cuts as

$$E(A) = \sum_{s=0}^{n-1} \sum_{t=0}^{n-1} \left[\left(\sum_{i=i_s}^{i_{s+1}} \sum_{j=j_t}^{j_{t+1}} p(i,j) \right) \left(\sum_{i=i_s}^{i_{s+1}} \sum_{j=j_t}^{j_{t+1}} 1 \right) \right] \tag{18}$$

where in the outer sums it is understood that the 0^{th} and n^{th} cuts correspond to the boundaries.

For simplicity we specialize to the subclass of distributions of hitting probabilities for which the i and j dependence factors: $p(i,j) = p(i)p(j)$. In this case Eq. (18) can be written as the product of quantities which depend separately on the positions of the i and j cuts :

$$E(A) = \left[\sum_{s=0}^{n-1} \left(\sum_{i=i_s}^{i_{s+1}} p(i) \right) \left(\sum_{i=i_s}^{i_{s+1}} 1 \right) \right] \times \left[\sum_{t=0}^{n-1} \left(\sum_{j=j_t}^{j_{t+1}} p(j) \right) \left(\sum_{j=j_t}^{j_{t+1}} 1 \right) \right]. \tag{19}$$

The optimal configuration minimizes $E(A)$ with respect to the position of the $2(n-1)$ cuts. The factorization allows us to solve for the positions of the i and j cuts separately. When the distribution $p(i,j)$ is centered at a point $i=j$, the i and j solutions are identical. When the distribution $p(i,j)$ is centered at the origin, the solution is symmetric around the origin.

We obtain an explicit numerical solution by minimizing the expected event size with respect to all possible placements of the cuts. Our result for an optimal grid subject to a Gaussian distribution of hits centered at the origin is illustrated in Fig. 3c (where the system size is taken to be relatively small to allow a visual comparison with the SOC state in Fig. 3a). Figure 1b illustrates analogous results for an asymmetric distribution with Gaussian tails, which is peaked at the upper left corner of the lattice. The corresponding distribution of event sizes is included in Fig. 2b. The distribution of event sizes for the symmetric case in a somewhat larger system is included in Fig. 4d. The cumulative distribution of events is reasonably well fit by a power law with $P_{\text{cum}}(A) \sim A^{-\gamma}$ with $\gamma \approx 3/2$.

Sharper estimates for the exponents can be obtained in the continuum limit, where we rescale the lattice into the unit interval ($x = i/N$, $y = j/N$) and take the number of lattice sites N to infinity. In the limit, the cuts become infinitesimally thin $d-1$ dimensional dividers between continuous connected regions of high density. We begin by solving the problem for $d=1$ since the solution to our grid problem factors into two one-dimensional problems. In each case, we adjust the positions of $(n-1)$ dividers to define n total regions, such that the minimum expected event size is obtained. Here the event size is associated with the length $L(\mathcal{R})$ of each of the regions.

To locate the positions of the cuts which yield the minimum expected size, we apply the variational method [26] separately to each bracketed term on the right hand side of Eq. (19). Determination of the stationary point with respect to the positions of each of the $(n-1)$ cuts yields an iterative solution for the cut positions:

$$\mathcal{P}(\mathcal{R}_i) + L(\mathcal{R}_i)p(x_i) - \mathcal{P}(\mathcal{R}_{i+1}) - L(\mathcal{R}_{i+1})p(x_i) = 0. \quad (20)$$

The cut positions beginning at the origin are obtained by solving Eq. (20) numerically. In Fig. 4b we illustrate $P(L)$ for cases in which $p(x)$ is described by Cauchy ($p(x) = \frac{\alpha}{\pi}(\alpha^2 + x^2)^{-1}$ with $\alpha = 1$), exponential ($p(x) = \sigma^{-1}\exp(-|x|/\sigma)$, with $\sigma = 10$), and Gaussian ($p(x) = (2\pi\sigma^2)^{-1/2}\exp(-x^2/2\sigma^2)$, with $\sigma = 15$) distributions. The parameters are chosen so that the optimal solution obtained from Eq. (20), involves a cut at the origin, followed 10 cuts in the half space ranging from $x \in [0, 10^4]$.

For the Gaussian and exponential cases, even on a logarithmic scale regions of small L are heavily clustered near the origin. For larger values of x consecutive region sizes grow rapidly with x, and the effect is most pronounced for the distributions in which the rate of decay of $p(x)$ is greatest. In the Gaussian case, the final region encompasses most of the system ($L_{10} = 9950$ out of the total length of 10^4, while the first nine regions are clustered within a total length of 50). The next value, L_{11}, is sufficiently large that it cannot be represented as a floating point number on most machines. For the Cauchy distribution, the lengths do not spread out on a logarithmic scale.

Like the more general case discussed Section II, the solution for the grid design yields power laws for a broad class of $p(x)$. Unlike the results in Section II where the scaling exponents were sensitive to the specific choice of $p(x)$, for this case we find that asymptotically $P(L) \sim 1/L$ for the Cauchy, exponential, and Gaussian distributions. In all three cases, the slope of $\log(P(L))$ vs. $\log(L)$ never gets steeper than -2.

A simple argument will help us see why the numerical observation that asymptotically $P(L) \sim 1/L$ is plausible. Note that in each case the decay of $p(x)$ is monotonic, so there are no repeated region sizes. Thus consecutive points in the distribution of event sizes $P(L)$ vs. L are obtained directly from consecutive terms in Eq. (20), namely $\mathcal{P}(\mathcal{R}_i)$ vs. $L(\mathcal{R}_i)$. If $P(L) \sim L^{-1}$ then the slope on a logarithmic plot:

$$\begin{aligned}\frac{\Delta\log(\mathcal{P})}{\Delta\log(L)} &= \frac{\log(\mathcal{P}(\mathcal{R}_{i+1})) - \log(\mathcal{P}(\mathcal{R}_i))}{\log(L(\mathcal{R}_{i+1})) - \log(L(\mathcal{R}_i))} \\ &= \frac{\log(\mathcal{P}(\mathcal{R}_{i+1})/p(x_i)) - \log(\mathcal{P}(\mathcal{R}_i)/p(x_i))}{\log(L(\mathcal{R}_{i+1})) - \log(L(\mathcal{R}_i))} \quad (21)\end{aligned}$$

will asymptotically approach -1. The second term in the denominator is asymptotically negligible compared to the first since the regions sizes are large and grow rapidly with increasing x. Combining this with Eq. (20) a slope of -1 is obtained as long as the first term in the numerator of (21) is negligible compared to the second. Asymptotically, we can extend the upper limit of the integral representation of $\mathcal{P}(\mathcal{R})$ in Eq. (17) to infinity. Then clearly $\mathcal{P}(\mathcal{R}_i) >> \mathcal{P}(\mathcal{R}_{i+1})$. If $p(x)$ decays too rapidly (e.g. double exponentially), the first term becomes negatively divergent when the logarithm is evaluated. However, this does not occur for distributions which the decay less sharp. Indeed, for the Cauchy, exponential, and Gaussian distributions we consider the first term in the numerator of Eq. (21) is negligible compared to the second, so that in each case asymptotically $P(L) \sim 1/L$. For the Gaussian and exponential cases the numerics blows up before we reach the asymptotic limit. For the Cauchy distribution, the fit to the asymptotic result is excellent.

The cumulative distributions $P_{\text{cum}}(L)$ are illustrated in Fig. 4c. These are obtained from Fig. 4b by summing probabilities of events of size greater than or equal to L. The solution for the $d = 2$ grid is obtained by overlapping the two one-dimensional solutions. The areas of the individual regions are given by $A(\mathcal{R}) = L_x(\mathcal{R})L_y(\mathcal{R})$ and the probabilities enclosed in each region is simply $\mathcal{P}(\mathcal{R}) = \mathcal{P}_x(\mathcal{R})\mathcal{P}_y(\mathcal{R})$. The results for power law, exponential, and Gaussian distributions of hitting probabilities are illustrated in Fig. 4d. In each case, the resulting distribution of event sizes exhibits a heavy tails, and is reasonably well fit by a power law. For comparison in Fig. 4d we include the results for the Gaussian case on the discrete lattice, numerically optimized with far fewer cuts. We obtain surprisingly good agreement with the continuum results for the exponent in the power law in spite of the sparse data and the finite grid effects which prevents us from obtaining an exact solution to Eq. (20) for the discrete lattice. Discrete numerical results are expected to converge exactly to the continuum case in the limit as $n, N \to \infty$ with $n/N \to 0$.

Finally, we emphasize that neither our choice to use a grid in the optimization problem nor our use of a factorizable distributions of hitting probabilities are required to obtain power laws tails for the distribution of events. We have verified that similar results are obtained for concentric circular and square regions, and for different choices of $p(i,j)$. The generality of our results suggests that heavy tails in the distribution of events follow generically from optimization of a design objective and minimizing hazards in the presence of resource constraints.

V: Evolution to the HOT State

Most systems in engineering and biology are not designed by global optimization, but instead evolve by exploring local variations on top of occasional structural changes. Biological evolution makes use of a genotype, which can be distinguished, at least abstractly, from the phenotype. In engineering the distinction is cleaner, as the design specifications exist completely independently of any specific physical instance of the design. In both cases, the genotype can evolve due to some form of natural selection on yield.

For both the primitive design and sophisticated grid design discussed in Section III, we can view the design parameters as the genotype and the resulting configuration as the phenotype. In the primitive design the density is the only design parameter. In the advanced design, the design parameter is the locations of the cuts.

By introducing a simple evolutionary algorithm on the parameters we can generalize the models so that they evolve to an optimal state for either the primitive or sophisticated design. The simplest scenario would involve a large ensemble of systems that evolve by natural selection based on yield. This is a trivial type of evolution, but it is obvious that such a brute force approach will be globally convergent in these special cases because the search space of cuts is highly structured. Interestingly, both cases evolve to a state which exhibits power law distributions, while all other aspects of the optimal state are determined by the design constraints. Even in the case of primitive design, the evolution proceeds by selecting states with high yield, and which differs from the internal mechanism by which SOC systems evolve to the critical point. With more design structure, systems will evolve to densities far above criticality.

Alternatively in the context of percolation, we can consider a local and incremental algorithm for generating configurations which is reminiscent of evolution by natural selection. We begin with an empty lattice, and occupy sites one at a time in a manner which maximizes expected yield at each step. We choose an asymmetric $p(i,j)$:

$$\begin{aligned} p(i,j) &= p(i)p(j) \\ p(x) &\propto 2^{-[(m_x+(x/N))/\sigma_x]^2} \end{aligned} \tag{22}$$

where $m_i = 1$, $\sigma_i = 0.4$, $m_j = 0.5$ and $\sigma_j = 0.2$, for which the algorithm is deterministic. We choose the tail of a Gaussian to dramatize that power laws emerge through design even when the external distribution is far from a power law. Otherwise Eq. (22) is chosen somewhat arbitrarily to avoid artificial symmetries in the HOT configurations.

Implementing this algorithm we obtain a sequence of configurations of monotonically increasing density, which passes through the critical density p_c unobstructed. Here p_c plays no special role. At much higher densities there is a maximum yield point followed by a drop in the yield. The yield curve $Y(\rho)$ is plotted in Fig. 2a for the $p(i,j)$ given in Eq. (22).

This optimization explores only a small fraction of the configurations at each density ρ. Specifically, $(1-\rho)N^2$ of the $\binom{N^2}{(1-\rho)N^2}$ possible configurations are searched. Nonetheless, yields above 0.9 are obtained on a 32×32 lattice, and in the thermodynamic limit the peak yield approaches the maximum value of unity. While the clusters are not perfectly regular, the configuration has a clear cellular pattern, consisting of compact regions enclosed by well defined barriers. As shown in Fig. 2b, the distribution of events $P(A)$ exhibits a power law tail when $p(i,j)$ is given by Eq. (22). This is the case for a broad class of $p(i,j)$, including Gaussian, exponential, and Cauchy.

Interestingly, in the tolerant regime our algorithm produces power law tails for a range of densities below the maximum yield, and without ever passing through a state that resembles the (fractal) critical state. This is illustrated in Figs. 1d and 2b where we plot the event size distribution $P(A)$ (lower of the "evolved" curves) for a density which lies below that associated with the peak yield. Note that this configuration has many clusters of unit size $A = 1$ in checkerboard patterns in the region of high $p(i,j)$ in the upper left corner.

VI: Contrasts Between Criticality and HOT

Our primary result is that the designed sand piles and percolation model produce power law distributions by a mechanism which is quite different from criticality. The fact that power laws are not a special feature associated with a single density in the HOT state is in sharp contrast to a traditional critical phenomena.

It is interesting to contrast the kind of universality we obtain for the HOT state with that of criticality. For critical points, the exponents which describe the power laws depend on a limited number of characteristics of a model: the dimensionality of the system, the dimensionality of the order parameter, and the range of the interactions. In the case of nonequilibrium systems, and particularly for SOC, the concept of universality is less clear. There are numerous examples of sand pile models in which a seemingly very minor change in the toppling rule results in a change in the values of the scaling exponents [22,27].

As discussed in Section II, for the HOT state we return to a case in which only a few factors influence the scaling exponent for the distribution of events. These include the exponent α which characterizes how the measure of size scales with the area impacted by an event, β which relates the area of an event to the resource density, and most importantly the tails of the distribution of perturbations $p(\mathbf{x})$. In this sense, many models of cascading failure yield the same scaling exponents, and thus may be said to fall into the same *optimality* class.

To further illustrate the differences we now consider quantities other than the distribution of events. For example, the fractal regions characteristic of events at criticality are replaced by regular, stylized, regions in the HOT state. Indeed our sophisticated designs are a highly simplified example of self-dissimilarity, in sharp contrast to the self-similarity of criticality. Although this concept has been suggested in the context of hierarchical systems, the basic notion that the system characteristics change dramatically and fundamentally when viewed on different scales clearly holds in our case. Put another way, renormalizing the sophisticated designs completely destroy their structure. While some statistics of the HOT state, such as time histories of repeated trials, may exhibit some self-similar characteristics simply because of the power law distribution, the connection with an underlying critical phenomenon and emergent large length scales which are central features in SOC are not present in the HOT state.

One of the most interesting differences arises when we consider the sensitivity of the HOT state to changes in the hitting probability density $p(i,j)$. In random systems, qualitatively and in most cases quantitatively similar results are obtained regardless of the probability density describing placements of the sparks or grains. In the BTW model a system which is driven at a single point produces a distribution of events which is essentially identical to the results obtained when the system is driven uniformly. In contrast, the HOT state is much more sensitive. The optimal design depends intrinsically on $p(i,j)$. Furthermore, if a system is designed for a particular choice of $p(i,j)$, and then is subject to a different density, the results for the event size distribution change dramatically.

This is illustrated in Figure 6, where we initialize the system in the optimal grid designed state for a Gaussian $p(i,j)$ centered at the origin, but then subject the system to a spatially uniform distribution of hits. The resulting event size distribution *increases* with the size of the event, where for the largest events $P(A) \sim A$. In this sense, random critical systems are much more generically robust than HOT systems with respect to unanticipated changes in the external conditions.

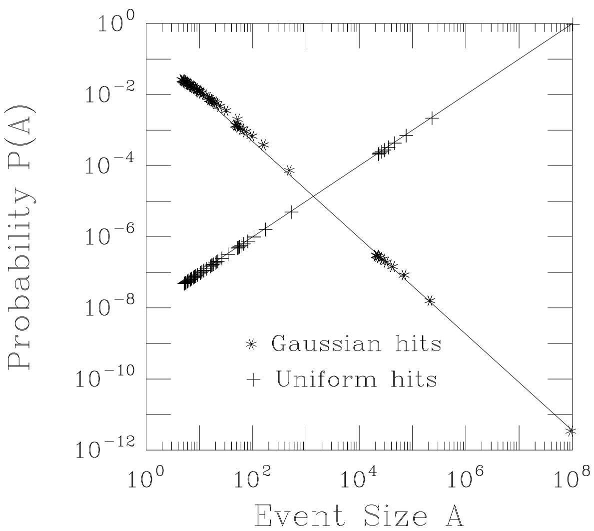

FIG. 6. The HOT state is highly sensitive to the distribution of hitting probabilities $p(i,j)$. Here we illustrate the probability $P(A)$ of an event of size A for the configuration designed for a Gaussian $p(i,j)$ in Fig. 4d. The points marked $\ast$ correspond to the results when the system is subject to the distribution of hits it was designed for (the results shown in Fig. 4d are obtained from these results by computing the cumulative number of events greater than or equal to A for each A). In contrast, $+$'s correspond to the case when the system is subject to a uniform distribution of hits. In this case the probability of large events exceeds the likelihood of small events.

Another sense in which the HOT state exhibits strong sensitivity relative to SOC is in terms of vulnerability to design flaws. A single flaw may allow an event to leak past the designed barrier. Furthermore, without incorporating a mechanism for repairing the system, repeated events gradually erode the barriers which leads to an overfrequency of large events that ultimately reduces the density to the critical point. This is illustrated in Fig. 7a for the case of a sand pile model with an initially uniform grid (similar results are obtained when the initial state is optimized for, e.g., a Gaussian).

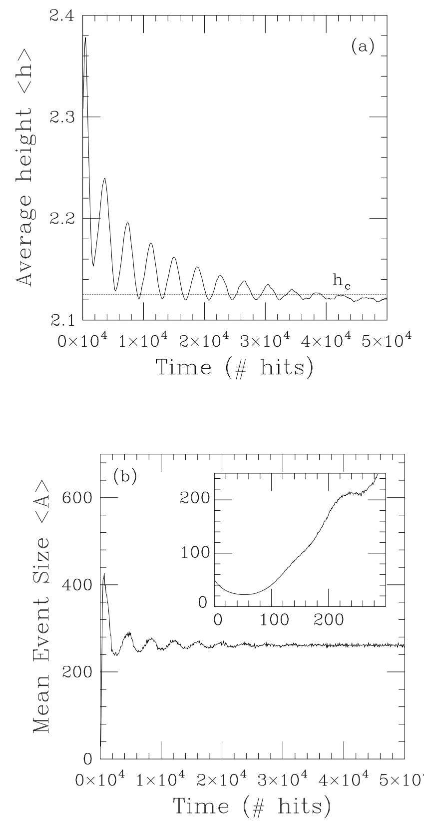

FIG. 7. Repeated events on the designed sand pile. Figure (a) illustrates the density vs. time. Initially the density oscillates– before the boundaries surrounding the center region have fully disintegrated, mass periodically accumulates. Eventually the system evolves back to the SOC state. In (b) we illustrate the corresponding mean event size vs. time. The mean event size initially decreases (shown on an expanded scale in the inset) as the grid contracts around the center n sites. These results are obtained on the discrete lattice for $N = 64$, initialized with 7 equally spaced vertical and horizontal cuts. The Gaussian distribution of hits is centered in the middle of the lattice, with $\sigma = 4$, and is computed as the average over 10^5 realizations. Results at small times converge rapidly, since each realization begins with the same initial state. We plot the mean over a large ensemble to obtain smoother results at long times.

While the HOT state is highly sensitive to unanticipated perturbations or flaws, additional constraints can be imposed on HOT designs to increase their robustness to any desired level, but at the cost of reduced performance. At the critical density, for example, it would be easy to design HOT states with small isolated clusters that would be highly robust to changes in probability distributions or flaws. Common strategies employed in biology and engineering to improve the system lifetime incorporate backup boundaries at additional cost (e.g. cuts which are more than one grid spacing in width). or mechanisms for the system to be repaired with regular maintenance. Engineers routinely add generic safety margins to protect against unanticipated uncertainties.

It is interesting to note that even large events on the designed sand pile do not immediately destroy the design structure when it is subject to repeated hits. When a grain is dropped directly on a cut, the height at that site increases but no avalanche occurs. When an avalanche is initiated within a rectangular domain the net effect is that the boundaries on all 4 sides step one site in towards the center of the box, and leave a residual site of reduced height at the previous corner points. All other sites return to their original height. Thus implementation of an elementary algorithm for repairing damage to the system should be straightforward.

Our observation that the net change associated with an avalanche in the grid design is simply to displace the boundaries one step towards the site that was hit suggests some degree of evolution towards the optimal state is intrinsic to the BTW algorithm. In Figure 7 we illustrate what happens when we begin with a regular grid of equally spaced cuts, and subject the system to repeated events using the BTW algorithm with hitting probabilities chosen from a Gaussian $p(i,j)$ centered in the middle of the lattice. We run a long sequence of repeated events without making repairs, and find that the mean event size initially decreases during a period in which the density is actually increasing (Fig. 7a), as the boundaries contract around the center of the lattice as illustrated in Fig. 7b. However, the designed sand pile never reaches the HOT state by this method. Repeated hits create sufficient flaws in the boundary that large events eventually return the system to the SOC state. However, as illustrated in the density plot (Fig. 7a), the transient period is extremely long.

VII: Conclusion

In summary, we have described a mechanism whereby design optimization in the presence of constraints and uncertainty naturally leads to heavy tailed distributions. Common features of the HOT state include (1) high efficiency, performance, and robustness to designed-for uncertainties, (2) hypersensitivity to design flaws and unanticipated perturbations, (3) nongeneric, specialized, structured configurations, and (4) power laws. We are not suggesting that HOT is the only alternative to SOC which yields power laws. In many cases, statistics alone may be responsible [28]. Furthermore, it seems likely that in some cases real systems may combine SOC or some other randomizing phenomenon with design in the process of mutation and selection as they evolve towards complex and efficient operating states.

An important consequence of the special features of the HOT state is the development of new sensitivities at each step along the path towards increasingly realistic models. Unlike criticality, where systems fall into broad universality classes which depend only on very general features, for HOT systems the details matter.

From a technological and environmental viewpoint, perhaps the most important feature of HOT states is the fact that the high performance and robustness of optimized designs with respect to the uncertainty for which they were designed, is accompanied by extreme sensitivity to additional uncertainty that is not included in the design. We considered changes to the hitting probabilities and flaws in the initial conditions, but other changes in the "rules" would have similar effects. In contrast, the SOC state performed relatively poorly, but was much less sensitive to changes in the rules.

This is one of the most important properties of complex biological and engineering systems that has no counterpart in physics, that complexity is driven by profound tradeoffs in robustness and uncertainty. Indeed there are fundamental limitations that can be viewed as "conservation principles" that may turn out to be as important as those due to matter, energy, entropy and information have been in the past [29].

We conclude with a brief discussion of two examples, one (the Internet) chosen from engineering, and one (ecosystems) chosen from biology. While both have been considered previously in the context of SOC/EOC, they clearly exhibit all the features associated with the HOT state. In discussing these examples, we will not attempt to provide a comprehensive review of the relevant literature, which is extensive in each case. We will simply illustrate (for an audience which is at least somewhat familiar with these disciplines) why these systems are good candidates for further investigations in the context of HOT. It is important to emphasize that our highly simplified models should not be taken seriously as prototypes for these particular systems. Instead, it is our intention to use toy models to illustrate several essential ingredients in "how nature works" which are absent in SOC. It is the general properties of HOT states, rather than the specifics of the percolation and sand pile models on the one hand, or internets or ecosystems on the other, which are common to a wide range of applications, and which therefore should be taken into account in the development of domain specific models.

VII.A HOT features of the Internet

We begin with the Internet which, as mentioned in the introduction, is an astonishingly complex system. Here we highlight a few issues that underscore the HOT features, including ubiquitous power law statistics. Computer networks are particularly attractive as a prototype system since a great deal of statistical data is available and experiments are relatively easy to perform, certainly compared with ecosystems. The history of the various types of networks that have been implemented also yields a rich source of examples. For example, familiar broadcast ethernet, but without collision and congestion control, would correspond to a (purely) hypothetical "random" network and would indeed exhibit congestion induced phase transitions at extremely low traffic densities. It is not hard to imagine that such a primitive and inefficient network could be made to operate in a state that might resemble

SOC/EOC.

In contrast, modern networks use routers and switches together with sophisticated control protocols to produce networks which are many orders of magnitude more efficient than if those routers, switches and protocols were removed. Thus the internal configuration is highly structured and specialized, and extremely robust to the main sources of uncertainty, which are due to user behavior and network component failure. The network is also hypersensitive to common-mode software bugs for which it is not designed, and thus has all the HOT features.

While the Internet, and computer systems more generally, have self-similar network traffic and ubiquitous power law statistics for everything from ftp and web file transfers to CPU usage [15,30], it remains somewhat controversial as to the origins of these effects and their significance for network design. It is widely agreed, however, that the "bursty" nature of network traffic requires, say, much large router buffers than would result from a more traditional queueing theory analysis [15]. A popular theory claims that "bursty" Internet traffic can be traced to power law distributions in web files [15,32]. Roughly speaking, this theory argues that large web file transfers due to heavy tails are streamed onto the network by TCP to produce long-term correlations, and thus burstiness and self-similarity in network traffic. This mechanisms seems to explain the burstiness on time scales of seconds to hours, that is, long compared to the round-trip packet times.

Tracing the origins of network burstiness to heavy-tailed web file distributions is an attractive starting point for understanding the power laws in a wide variety of measurements since it is consistent with the observation that the (long-time) burstiness is independent of congestion level. Recall that, based on the evolutionary model (Section V), we have identified power laws at all densities above criticality as a distinction between HOT and criticality. While this theory explains network burstiness in terms of heavy tails in web files, so far there is no accepted explanation for the heavy tailed web file distributions, despite enormous statistical evidence for them [32–35].

We suspect that the power laws in web file distributions may arise via HOT. That is, HOT features may extend not only to the network but to the web sites themselves. High-volume commercial web sites are constantly tuned for high throughput, and thus we can explore what properties might be consequences of such design. A simple model for this would be to assume that the "document" making up a web site is partitioned into files to minimize the expected sizes of file transmissions. Users exhibit widely varying levels of interest in the document, so that an "optimized " web site would have smaller files for high hit portions of the document. To make the connection more precise, suppose that we model user interest as a probability distribution $p_u(x)$ where x is the location within the document that the user would like to examine. Real web documents, of course, have a great deal of a priori structure, but we will make the the highly idealized assumption that the document itself is just a single contiguous object. Also, real users interact in complex ways with the structure of the document. Thus a model that assumes the user is interested in a single location in an unstructured document is extremely simplified, but allows us to use the results in Section IV.

An abstract web design problem would then correspond to partitioning the document into N files such that the expected file transfer is minimized. Because a hit on a file causes the entire file to be transferred, the expected transfer size $E(\mathcal{S})$ is given by a sum over the files i of the product of the probability of the file $\mathcal{P}(i)$, obtained from the probability $p_u(x)$ that x will be in file i, and the size of the file $\mathcal{S}(i)$:

$$E(\mathcal{S}) = \sum_i \mathcal{P}(i)\mathcal{S}(i). \tag{23}$$

Minimizing $E(\mathcal{S})$ corresponds to exactly the optimization problem we solved in Section IV for the grid design. In that case variational methods led to Eq. (20) for the positions of the cuts in one dimension, which in this case correspond to cuts in the document, breaking it up into a set of individual files. Asymptotically in Section IV we found that for a broad class of probability distributions for the hits we indeed obtain heavy tails. Superficially, the plots in Figure 4c for the resulting cumulative distributions do resemble those for web sites, but this shouldn't be taken too seriously as it is not a statistically precise comparison.

This view of website design is so idealized that it may not explain in any detail why real web sites have power law distributions. The assumption of a homogeneous document is particularly suspect, and intrinsic heterogeneity and hierarchy in the original document itself may be more important to the website layout than user interest. Also, users typically browse a website in a sequence that reflects the website's structure, and thus we are exploring models with more realistic structure. However, given how robust the HOT mechanism for producing heavy tails is, we expect that many different design elements could contribute in different settings, but all would yield the same effective network behavior. We hope that this approach may begin to demystify some of the discussion, since it shows that the observed power laws, including even (roughly) the exponents, are at least consistent with the web sites being designed. The constant tweeking of high volume commercial web sites to maximize throughput might yield an adaptive process which is a reasonable approximation to HOT. Further research in this direction, particularly with richer models for web documents and user interest, will be needed to evaluate the significance of our speculations.

VII.B HOT features of Ecological Systems

Finally, we move to ecosystems. In comparison to the Internet, here the analogy while suggestive is much less precise. For the Internet, we have access to a great deal of statistical information as well as all the details of how the system is designed. From this we are beginning to develop a case for HOT at the level of the file distributions on web sites, as discussed above, as well as the network as a whole. We are suspicious that a similar story may apply to ecosystems, but it is necessarily more speculative because we have a less complete understanding of the details. In the environmental literature, the definition of what is meant by "ecosystem" is in itself a topic of debate, and determining precisely how concepts such as "optimization," "yield," and "events" might play a role in the interactions between species is much more ambiguous. Nonetheless, modeling population dynamics [36] play a central role in environmental science. Furthermore, there is increasing evidence that the widespread observations of heavy tailed distributions arises as a consequence of the dynamical response of coupled populations to external disturbances [37]. In the case of environmental policy, there are fundamental distinctions between the implications of SOC/EOC and HOT.

It has been argued, principally by physicists, that ecosystems are in a critical state because the distribution of sizes of extinction events, as deduced from the fossil record, is characterized by a power law [38]. This fact has motivated the EOC based Kauffman–Johnsen model [39], which describes the evolution of coupled fitness landscapes, and the Bak-Sneppen model [40], which is a simple SOC model of a set of species evolving on a fitness landscape. However, there is an ongoing debate as to whether the SOC/EOC models capture the essential features of real environmental systems. The alternative perspective offered more typically by biologists and ecosystem specialists exhibits many features of HOT. Below we summarize a few key results in environmental studies which support this point of view.

Our investigation of the primitive (random) and sophisticated designs in percolation and sand pile models has direct parallels in studies of the role of increased complexity and structure in ecosystems. For ecosystems, the analog of moving towards higher densities is associated with increasing the number of organisms and/or increasing the number of species, which is referred to "increasing complexity" in the ecology literature. The early and influential work of Robert May [41] suggested that high density states (high levels of complexity in ecosystems) are not stable– in simple models increased population and differentiation eventually leads to a bifurcation analogous to the percolation transition in the random system. However, according to a recent review by Polis [42], "it was clear to empiricists and some theoreticians that natural systems are quite complex. In any one system, a great diversity of species is connected through many different interactions," in contradiction to May's conclusions that increasing complexity will eventually cause ecological systems to exhibit strong fluctuations and "fall apart."

More recent work by McCann, Hastings, and Huxel [43] has shown that increased density (i.e. complexity) tends to stabilize an ecosystem, damping out fluctuations and preventing the loss of species. Their work is based on models with a more accurate representation of the biology, and leads to systems which stabilize at higher densities, in a manner which is qualitatively similar to the way in which our sophisticated design in the evolutionary model (Section V) passes unobstructed through the critical point associated with a random system to reach a structured high density state.

Additional evidence for the critical importance of evolved structure in ecosystems is obtained in the study of food webs [44]. In simple randomized models high densities (i.e. complexities) destabilize food webs in a manner which again parallels the falloff in yield which we observe in random systems for densities which exceed the critical point. In numerical and laboratory studies, randomly assembled but plausible food webs typically break down [45,46]. The results of these studies are contradictory to observations of food webs in nature, which are composed of large numbers of interacting species. However, real food webs are not randomly constructed. In model studies, incorporating design features such as adjusting the distribution of interaction strengths to maximize survival and introducing redundancy, lead to the persistence of webs at more realistic densities, as well as optimized configurations which are consistent with observed variabilities in interaction strengths in natural systems.

Food web studies also illustrate the hypersensitivity of ecosystems to changes that the system was not designed to handle [44]. If food webs were in an SOC/EOC state then their complexity would be generic and robust to rearrangements and the introduction of new species. However, in real ecosystems the introduction of one weedy species, such as an exotic plant or animal species, often leads to catastrophic consequences [47].

Finally, returning to the extinction patterns in the fossil record, there is some controversy over whether the data is actually well described by a power law [48]. What is clear is that there have been mass extinctions and that the distribution of events is at least plausibly a power law. Almost all species that have existed are extinct (0.1% of all recorded species currently persist), and the average lifetime of a species is of order a million years, though the distribution of lifetimes also has heavy tails.

There is a long running debate in paleontology about the relative roles of random versus deterministic effects in extinctions. This may be a false dichotomy. If ecological systems and populations are HOT, then it is the interaction between the unpredictable external perturbations and the structured state of the system which is crucial.

Among paleontologists there is general consensus that extinction vulnerability has some systematic features, consistent with HOT states [49]. Specifically, organisms evolve to maximize their survival within the existing environment, and thus become vulnerable to rare events. For example, large size, and high specialization may yield a temporary advantage, but creates vulnerability to, say, meteor impacts.

Thus while extinctions may be triggered by exogenous events, the pattern of extinctions for a given disturbance follows a fairly structured, deterministic, and even a predictable process. In a large event it is the most complex, terrestrial organisms which are at greatest risk.

Island species are more vulnerable than continental specials. Tropical species are more vulnerable than nontropical species. Within a habitat, specialization offers short-term benefits. However, evolution towards more specialized states inevitably ignores rare events, even if they are catastrophic. As a consequence, specialization consistently correlates with extinction risk in large extinctions. This suggests that the system can be viewed as HOT on two interdependent levels. The fact that the overall extinction patterns are highly structured suggests that ecosystems as a whole may be viewed as HOT, while the fact that the most specialized organisms are most vulnerable suggests that the evolution of individual species may lead them towards increasingly HOT states.

There is much at stake in this debate. If ecosystems are in a SOC/EOC state, then observations of massive species extinctions and global warming could be attributed to the natural behavior of the system. In this scenario, large fluctuations emerge and recede as a natural consequence of the internal dynamics, and would not be attributed to man made causes. This would support a policy in which humans could be relatively cavalier about their interactions with the environment, because the system would be fluctuating as observed regardless of our behavior. Alternately, if ecosystems are in a HOT state then we expect the system to be robust, yet fragile. Heavy tailed distributions are expected, but the system is also hypersensitive to new perturbations that were not part of the evolutionary history.

In terms of policy, this supports a strategy of cautious environmental perturbation. Polis writes [42], "From a policy point of view, the understanding that complexity is vital to the integrity and stability of natural systems allows ecologists to argue, more coherently, why we must preserve the diverse elements and species that coexist in a healthy, sustainable and well-functioning ecological community. " If environmental systems are HOT, then the burden of proof in the ecological debate must shift from a policy of "waiting for the science" to confirm negative effects such as ozone depletion or global warming, to a policy which requires substantial scientific investigation of whether the perturbed system is robust to proposed changes before they are introduced.

Acknowledgements : The authors would like to thank the many people who contributed ideas and comments to this paper. We would particularly like to thank Fernando Paganini for suggestions about Section II, and Mani Chandy, Mor Harchol-Balter, Walter Willinger, Vern Paxson, and Sanjay Lall for discussions about the Internet. This work was supported by the David and Lucile Packard Foundation, NSF Grants No. DMR-9212396 and DMR-9813752, and a DOD MURI grant for "Mathematical Infrastructure for Robust Virtual Engineering."

---

[1] N. Goldenfeld and L. Kadanoff, *Science* **284**, 87-89 (1999).
[2] G. Nicolis and I. Prigogine, *Self Organization in Nonequilibrium Systems* (Wiley, New York, 1977).
[3] P. Bak, C. Tang, K. Wiesenfeld, *Phys. Rev. Lett.* **59**, 381 (1987).
[4] See, e.g., P. Bak, *How Nature Works: The Science of Self-Organized Criticality,* (Copernicus, New York, 1996).
[5] S. A. Kauffman, *The Origins of Order: Self-Organization and Selection in Evolution,* (Oxford Univ. Press, New York, 1993).
[6] B. Mandelbrot, *The Fractal Geometry of Nature* (Freeman, New York, 1983).
[7] C. C. Barton and P. R. Lapointe, *Fractals in the Earth Sciences,* (Plenum Press, New York, 1994).
[8] M. Schroeder, *Fractals, Chaos, and Power Laws,* (Freeman, New York, 1991).
[9] R. Lewin, *Complexity - Life at the Edge of Chaos,* (Macmillan, New York, 1992).
[10] M. M. Waldrop, *Complexity: The Emerging Science at the Edge of Order and Chaos,* (Simon & Schuster, New York, 1992).
[11] B. Goodwin, *How the Leopard Changed Its Spots - The Evolution of Complexity,* (Charles Schribner's Sons, New York, 1994).
[12] L. Smolin, *The Life of the Cosmos,* (Oxford Univ. Press, New York, 1997).
[13] A. Gore, *Earth in the Balance* (Houston Mifflin, Boston, 1992).
[14] W. E. Leland, M.S. Taqqu, W. Willinger, and D. V. Wilson, *ACM/SIGCOMM Computer Communications Review*, **23**, 183 (1993).
[15] W. Willinger and V. Paxson, *Notices of the AMS* **45**, 961 (1998).
[16] W. Stallings, *High-Speed Networks : Tcp/Ip and Atm Design Principles* , (Prentice Hall, New York, 1997).
[17] See ,e.g., *Sierra Nevada Ecosystems Project,* Congressional Report, (Regents of the University of California, 1996). Available on line at http://ceres.ca.gov/snep
[18] See, e.g., the Weyerhaeuser plan at http://www.link.ca/20yearplan/index.htm
[19] P. Bak, K. Chen, and C. Tang *Phys. Lett A*, **147**, 290 (1990); B. Drossel and F. Schwabl *Phys. Rev. Lett.* **69**, 1629 (1992).
[20] See, e.g., B.Alberts, D. Bray, A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter, *Essential Cell Biology*, (Garland, New York, 1998).
[21] D. Stauffer, *Introduction to Percolation Theory,* (Taylor, London 1985).
[22] J. M. Carlson, J. T. Chayes, E. R. Grannan, and G. H. Swindle, *Phys. Rev. Lett.* **65**, 2547 (1990); J. M. Carlson, E. R. Grannan, C. Singh, and G. H. Swindle, *Phys. Rev. E* **48**, 688 (1993); J.M. Carlson and G.H. Swindle, *Proc. Nat. Acad. Sci.* **92** 6712-6719, 1995.
J. M. Carlson, E. R. Grannan, C. Singh, and G. H. Swindle, *Phys. Rev. E* **48**, 688 (1993).
[23] O. Narayan and A. A. Middleton, *Phys. Rev. B*, **49**, 244 (1994).
[24] M. Paczuski, S. Maslov, P. Bak, *Phys. Rev. E* **53**, 414 (1996).
[25] There are no long range correlations in the BTW model. However, there are some known rules based on local

properties which define the subset of allowed configurations. These are described in D. Dhar, *Phys. Rev. Lett.* **64**, 1613 (1990).

[26] Strictly speaking, the variational calculation performed in Section II does not apply to this case. The fixed number of cuts, when expressed in terms of a constraint equation analogous to (2) involve a $C(X)$ which is not differentiable.

[27] L. Kadanoff, S. Nagel, L. Wu, and S. Zhou, *Phys. Rev. A* **39**, 6524 (1989).

[28] See, e.g., A. Janicki and A. Weron, *Simulation and Chaotic Behavior of α-Stable Stochastic Processes* (Marcel Dekker, Inc. New York, 1994).

[29] M. M. Seron, J. H. Braslavsky, and G. C. Goodwin, *Fundamental Limitations in Filtering and Control*, (Springer Verlag, New York, 1997).

[30] M. Harchol-Balter, M. Crovella, C. Murta, *Lecture Notes in Computer Science, No. 1469. Computer Performance Evaluation. Modelling Techniques and Tools*, (Edited by: R. Puigjaner, N. Savino, B. Serra, Berlin, Germany: Springer-Verlag), p. 231-42 (1998).

[31] Harchol-Balter, M.; Downey, A.B. *ACM Transactions on Computer Systems* **15**, 253 (1997); *Performance Evaluation Review* **24**, 13 (1996).

[32] M. E. Crovella and A. Bestavros, *IEEE/ACM Transactions on Networking* 5(6):835–846 (1997).

[33] M. E. Crovella, M. S. Taqqu, and A. Bestavros, *A Practical Guide To Heavy Tails*, 1–23. (Chapman & Hall, New York, 1998).

[34] G. Irlam. Unix file size survey 1993. Available at http://www.base.com/gordoni /ufs93.html (1994).

[35] V. Paxson and S. Floyd, *IEEE/ACM Transactions on Networking*, 226–244, (1995).

[36] J. Roughgarden, *Primer of Ecological Theory*, (Prentice Hall, NJ 1998).

[37] C. Zimmer, *Science* **284**, 83-86 (1999).

[38] D. M. Raup, *Science, Wash.* **231**, 1528-1533 (1986).

[39] S. Kauffman and J. Johnsen, *J. Theor. Bio.* **128**, 11-45 (1991).

[40] P. Bak and K. Sneppen, *Phys. Rev. Lett.* **71**, 4083-4086 (1993).

[41] May, R, *Stability and Complexity in Model Ecosystems* (Princeton Univ. Press, 1973).

[42] Polis, G., *Nature* **395**, 744 - 745 (1998).

[43] McCann, K., Hastings, A. and Huxel, G. R. *Nature* **395**, 794-798 (1998).

[44] I. Hanski, *Nature* **390**, 440-441 (1997).

[45] Pimm, S. L. *Ecology* **61**, 219-225 (1980).

[46] Morin, P. J. and Lawler, S. P. *Annu. Rev. Ecol. Syst.* **26**, 505-529 (1995).

[47] Y. Baskin, *BioScience* **48**, 788-794 (1998).

[48] J. W. Kirchner and A. Weil, *Nature* **395**, 337-338 (1998).

[49] Michael L. McKinney, *Annu. Rev. Ecol. Syst.* 28:495-516 (1997).