Three mechanisms for power laws on the Cayley tree

Ted Brookings and J. M. Carlson
Department of Physics, University of California, Santa Barbara, California 93106, USA

John Doyle
Control and Dynamical Systems, California Institute of Technology, Pasadena, California 91125, USA
(Received 5 August 2005; published 16 November 2005)

We compare preferential growth, critical phase transitions, and highly optimized tolerance (HOT) as mechanisms for generating power laws in the familiar and analytically tractable context of lattice percolation and forest fire models on the Cayley tree. All three mechanisms have been widely discussed in the context of complexity in natural and technological systems. This parallel study enables direct comparison of the mechanisms and associated lattice solutions. Criticality fits most naturally into the category of random processes, where power laws are a consequence of fluctuations in an ensemble with no intrinsic scale. The power laws in preferential growth can be understood in the context of competing exponential growth and decay processes. HOT generalizes this functional mechanism involving exponentials of exponentials to a broader class of nonexponential functions, which arise from optimization.

DOI: 10.1103/PhysRevE.72.056120 PACS number(s): 89.75.−k, 05.50.+q, 64.60.Ak

I. INTRODUCTION

Power laws in cumulative event frequency $P(\geqslant s)$ vs size s, $P(\geqslant s) \sim s^{-\alpha}$, arise commonly in observed data for a wide range of phenomena including natural, social, and technological disasters [1,2]. Power law statistics also arise frequently in models, and are often cited as evidence supporting different mechanisms for complexity. These include (i) criticality [3,4] and self-organized criticality (SOC) [5,6], (ii) preferential growth [7–9] and preferential attachment [10–12], and (iii) highly optimized tolerance (HOT) [13–21]. All three of these mechanisms have been discussed recently, and in most cases separately, in the context of a variety of different applications where power laws are observed, including biology [7], ecology [22–24], and the Internet [25]. From a purely mathematical point of view, power laws have strong statistical invariance properties, and are seen to arise from a variety of functional mechanisms (see Appendix A). Thus their ubiquity is not surprising. Power laws alone are not conclusive evidence for any particular mechanism. This mandates a more detailed, parallel examination of the assumptions that underly different mechanisms, and the corresponding mathematical solutions that lead to power laws, in order to draw distinctions between the mechanisms.

It is the purpose of this paper to illustrate the origins of the power laws in criticality, preferential growth, and HOT in the context of a single, analytically tractable lattice model: static site percolation on a Cayley tree. The Cayley tree (Sec. II) is a lattice that branches from a central site, but contains no loops. In percolation and forest fire models the branching nodes are sites, which can be occupied (trees) or vacant (fire breaks). The Cayley tree is a useful template for parallel, mathematical studies because analytical calculations are relatively straightforward. Presenting the mechanisms in a unified format concretely illustrates the intrinsic differences between the assumptions made in the different models, and highlights the consequences in terms of lattice configurations and event size statistics.

In statistical physics, it is well known that the size distribution of connected clusters of occupied sites for randomly generated lattice configurations exhibits a power law if and only if the density of occupied sites takes a specific critical value [26]. However, another mechanism for power laws on Cayley trees occurs for random lattices over a wide range of densities above the critical point in the boundary sizes at large lattice radii of the connected clusters. This can be viewed as a special case of preferential growth, an early model for bacteria population growth subject to mutations. HOT also produces power laws at high densities, but in nonrandom lattices as the result of optimization. HOT incorporates deliberate design or evolution in a manner that is motivated by engineering or biology. Unlike the other mechanisms, HOT configurations correspond to a set of measure zero in the ensemble of all possible configurations at a given density, and reflect robustness tradeoffs associated with a spectrum of external perturbations.

The remainder of this paper is organized as follows. In Sec. II we describe the Cayley tree and the Bethe lattice, which is an infinite Cayley tree defined in a manner that excludes boundary sites. (The Bethe lattice is used in the analytical calculations for percolation.) The lattice is a natural template for models of bacterial growth, because the branching structure can be associated with temporal evolution of the population through clonal reproduction. Alternatively, it has a natural interpretation as a quasi-infinite-dimensional spatial lattice in the context of criticality and HOT. In Secs. III–V we define and solve each of the models—preferential growth, criticality, and HOT—on the Cayley tree (Bethe lattice), with special attention to lattice configurations and derivations of the power laws. Preferential growth and criticality are described by essentially identical models (in both cases the system is described by the ensemble of configuration in which sites are occupied randomly), but exhibit power laws at different densities. HOT corresponds to a particular (optimized) high density configuration. In Sec. VI we compare the models directly in terms of

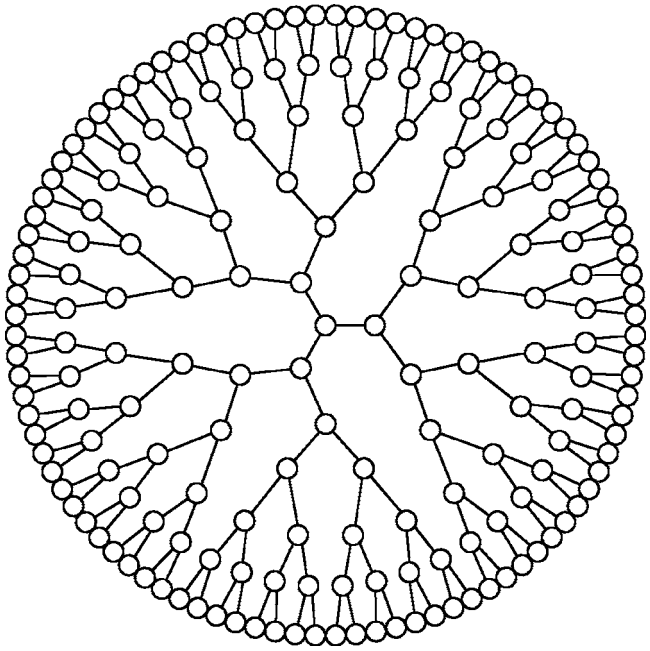

FIG. 1. A Cayley tree with $Z=3$, and seven "levels."

their scaling properties, spatial correlations, and robustness. We conclude in Sec. VII with a discussion of the models in the broader context of statistical and functional mechanisms for power laws. On the Cayley tree, both preferential growth and HOT (for an exponential distribution of disturbances) are examples in which power laws arise from competition between exponential growth, and exponential decay, a simple and versatile functional mechanism for generating power laws [33]. When the HOT model is solved numerically for different (nonexponential) disturbance patterns, the "exponentials of exponentials" mechanism appears to generalize to a broader class of nonexponential functions.

II. THE CAYLEY TREE

A *Cayley tree* with *coordination number Z*, is a branching graph where each node (unless it is on the edge) is connected to Z nodes with no loops. Figure 1 illustrates an example for $Z=3$, with seven levels branching from a central site.

A *Bethe lattice* is an infinite Cayley tree. The thermodynamic (infinite) limit is taken in a manner which eliminates the boundary effects, which are substantial, because the boundary corresponds to finite fraction of the Cayley tree (see below). In this paper, the Bethe lattice is used for analytical calculations associated with the percolation phase transition, a property associated with the interior, or bulk, of the lattice. All other calculations and all numerical results are performed on the Cayley tree.

A Bethe lattice is constructed by first considering an interior section far from the edge, of a large (but finite) Cayley tree. The thermodynamic limit is defined taking both the boundary of the Cayley tree and the interior section to infinity, so that the Cayley tree boundary goes to infinity much faster than the edge of the interior section. In this way, edge effects are formally eliminated for calculations on the Bethe lattice.

The branching structure of the Cayley tree (Bethe lattice) can be viewed as a discrete temporal evolution, as in preferential growth, or as a graph representing spatial connectivity, as in criticality and HOT. In the spatial case, the absence of loops in the Cayley tree (Bethe lattice) leads to solutions that are formally analogous to those obtained in infinite dimensions.

Imagine embedding a Bethe lattice in, say, a d-dimensional hypercubic lattice. This can only be done in the limit $d\to\infty$, because of the absence of loops. On an ordinary finite-dimensional lattice, loops become less important as dimension increases; and statistically cease to contribute in the limit of infinite dimensions. Due to the infinite-dimensional character of the Bethe lattice, mean field theory is exact. For this reason, the Bethe lattice is considered a standard template for mean field calculations. Furthermore, in many cases the absence of loops makes calculations on the Cayley tree or Bethe lattice tractable.

The analogy with infinite dimensions also arises in scaling relations between the $(d-1)$-dimensional surface area (A) and d-dimensional volume (V) of compact clusters, of characteristic length L on the lattice. To see this first consider a hypercubic lattice of length L on each side. The surface area of a hypercubic lattice, A_H, scales as

$$A_H \sim L^{d-1}, \tag{1}$$

and the volume V_H scales as

$$V_H \sim L^d. \tag{2}$$

Eliminating the L dependence leads to

$$A_H \sim V_H^{(d-1)/d}. \tag{3}$$

Next, consider a Cayley tree with N levels extending out from a central site (the "origin"). The corresponding asymptotic relationships (for large N) for surface area A_C and volume V_C are

$$A_C \sim Z(Z-1)^{N-1} \tag{4}$$

and

$$V_C \sim Z(Z-1)^N. \tag{5}$$

Hence, for the Cayley tree

$$A_C \sim V_C^{(N-1)/N}. \tag{6}$$

In the limit $N\to\infty$, surface area scales linearly with volume: $A_C/V_C \sim (Z-2)/(Z-1)$, i.e., asymptotically A_C scales linearly with V_C on the Cayley tree, which matches the previous result for the hypercubic lattice [Eq. (3)] in the limit $d\to\infty$. Note that this also implies that in the limit of large lattices, the boundary sites comprise a fraction $(Z-2)/(Z-1)$ of the total sites on the Cayley tree.

III. PREFERENTIAL GROWTH

Preferential growth is an early and well established quantitative model for population growth in biology [7,22]. It is motivated by clonal cell division in bacterial populations. Each new generation consists of clones of the previous, par-

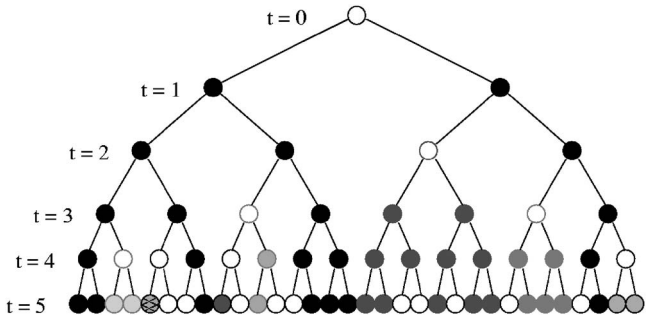

FIG. 2. Preferential growth with $Z=3$ and six levels or generations, and $\rho=0.75$. Empty circles are mutants, and filled circles are unmutated clone descendants. At the end, the largest strain size is 7 (the original strain).

ent generation, subject to a small possibility of random mutation (when a mutation occurs, we say it produces a new *strain*). Here *preferential* growth refers to the fact that large strains (i.e., strains that have a large population) in a given generation preferentially tend to flourish, and thus remain a large strain in the next generation. This is a similar mechanism (though not mathematically identical) to the preferential attachment mechanism which has received a great deal of attention recently, and which leads to the so called scale-free networks [11,27–32]. In preferential attachment, networks are grown in a manner that favors links between new nodes and sites which already have high connectivity (i.e., node degree) in the network. The resulting network exhibits power law distributions in the number of links entering and/or exiting a given node, but is different from the models studied here on the Cayley tree, because preferential attachment generates networks with loops.

Preferential growth leads to power laws in the size distribution of living strain sizes [33]. While originally studied in the context of continuous ordinary differential equations (ODE's) by Mandelbrot and others [7], here we define the model on a single branch of the Cayley tree. The Cayley tree is in fact a convenient and natural template for the model. On the Cayley tree each node is viewed as a single organism. This avoids issues associated with fractional cell counts, which inevitably arise in the ODE formulation. The forward branching structure of the Cayley tree is analogous to cell division, corresponding to discrete time steps, separating each offspring from the previous parent generation. Figure 2 can be viewed as a "family tree" tracing the lineage of a single parent cell, which is created at time, or generation, $t=0$. The boundary sites on the bottom represent the living cells in the current generation, $t=5$. Absent from this model are effects associated with limited resources, spatial constraints, and fluctuations in growth and reproduction times. Thus the preferential growth model presented here is a primitive model of bacterial growth, which has the primary advantage of capturing a mechanism for power laws within a framework involving minimal input ingredients.

Evolution of the population is defined on the Cayley tree as follows. We begin with a single parent cell at time $t=0$, corresponding to a node on the lattice. The colony grows in increments, marked by integer values of t, corresponding to each subsequent generation of offspring. At time t, a new generation is created, consisting of $(Z-1)$ offspring from each parent cell which existed at time $(t-1)$. With probability ρ a given offspring is an exact clone of the parent cell. With probability $(1-\rho)$ a given offspring is a mutant. Mutants are always considered to be different from any other cell type that is present in the colony, and thus each mutation starts a new strain of cell types in the colony.

This process is represented graphically in Fig. 2. The color coding is defined so that white nodes correspond to mutants, and colored nodes correspond to identical clones. Different colors correspond to different strains, arising from distinct mutation events. Identical colors represent genetically identical offspring that are part of the same strain, which all arise through clonal reproduction from a common parent. For example, the initial parent cell at $t=0$ is white, and thus corresponds to a new mutation. (We can think of this tree as a branch pruned from some larger lattice.) The two offspring that arise at $t=1$ are both identical clones of the parent cell, and are colored black. At $t=2$ three of the four offspring are also identical clones of the parent cells, and are thus also colored black. However, one of the cells is a mutant, colored white, which generates a new strain in the population. In the next generation $(t=3)$ we see that offspring of this new cell type will be colored dark gray, to distinguish them from the black colored offspring of the original parent cell. The process continues in this manner. At a given time t the size distribution of living strains is determined by counting the number of cells that have the same color (and thus can be traced back through the lineage to a common mutant parent cell), with the exception of the white colored cells, which are all new mutants, and thus distinct. At $t=5$ in Fig. 2 the largest strain is the one colored black (which contains seven total clones), which is descended from the original parent cell at $t=0$. The smallest strains each have only one cell, corresponding to new mutants (there are ten of these).

As long as the probability ρ that an offspring cell is identical to the parent cell is large, most strain sizes grow exponentially with time. What is large enough? As we will see shortly, this system is closely connected to percolation, where time t is replaced by a radial spatial index marking the distance from the origin (the original parent cell in preferential growth), exact clones (probability ρ) correspond to occupied sites (with no special distinction associated with different colors), and mutants [probability $(1-\rho)$] correspond to vacant sites. In fact, this gives an exact mapping between preferential growth and percolation. It is simply the quantities of interest that change. Percolation has a critical density, marking the onset of finite probability that the origin is connected to the boundary in the limit of infinite lattice size, and the average size of connected clusters diverges. Returning to the case of preferential growth, "large enough" for exponential growth of most strain sizes corresponds to ρ being large compared to that critical point in percolation, which is given by $\rho_c=1/(Z-1)$.

In preferential growth, the quantity of interest is the probability $P(\geqslant s)$ that the size of a randomly selected strain is at least s [i.e., selecting a strain at random from those which exist after a long time, $P(\geqslant s)$ is the probability there are at

least s identical clones in that strain]. If ρ is sufficiently large compared to the percolation critical point ρ_c, $P(\geqslant s)$ is a power law. We show this using a simple heuristic argument.

Define the generation of the original node to be $t=0$. Then at generation t, there are $(Z-1)^t$ total nodes. On average $(1-\rho)(Z-1)^t$ of these nodes will correspond to mutations relative to the corresponding parent node. Therefore, after t generations, the probability P_b (subscript b denotes "born") that a given strain was created at generation t_b is given by

$$P_b(t_b) = \frac{(1-\rho)(Z-1)^{t_b}}{\sum_{j=1}^{t} (1-\rho)(Z-1)^j}. \tag{7}$$

In fact, we are more interested in the probability that a strain was created at or before t_b, as we want the cumulative probability distributions (and as we show below, earlier creation times correspond to larger sizes). Thus, the probability that the strain was created at or before t_b is

$$P_b(\leqslant t_b) = \frac{\sum_{k=1}^{t_b} (1-\rho)(Z-1)^k}{\sum_{j=1}^{t} (1-\rho)(Z-1)^j}. \tag{8}$$

Computing the sums, this becomes

$$P_b(\leqslant t_b) = \frac{(Z-1)^{t_b}-1}{(Z-1)^t-1}. \tag{9}$$

Next assume that the strain arising from a new mutation grows *deterministically* once it is created (this is an approximation, which we test and verify numerically below). With each successive generation, the expected (i.e., average) value of the number of identical clones grows like

$$s(t,t_b) = s(t-1,t_b)(Z-1)\rho, \tag{10}$$

so given that each strain starts with size 1 when $t=t_b$, we have

$$s(t,t_b) = [(Z-1)\rho]^{t-t_b}. \tag{11}$$

For Eq. (11) to describe *growth*, $(Z-1)\rho$ must be greater than unity ($\rho > \rho_c$). In the case where $\rho \leqslant \rho_c$, strains do occasionally grow, but this is due to stochastic deviations from the mean (shrinking) behavior. In that case, this argument breaks down and a power law is not observed.

As stated earlier, here we see smaller t_b corresponds to larger size. Furthermore, we can invert Eq. (11) to obtain the time of creation,

$$t_b(s,t) = t - \frac{\ln s}{\ln[(Z-1)\rho]}. \tag{12}$$

From this we obtain the finite-time distribution of strain sizes $P(\geqslant s,t) = P_b(\leqslant t_b(s,t))$:

$$P(\geqslant s,t) = \frac{(Z-1)^t s^{-[\ln(Z-1)]/\ln[(Z-1)\rho]} - 1}{(Z-1)^t - 1}. \tag{13}$$

In the limit $t \to \infty$ the t dependence in Eq. (13) drops out, yielding the strain size distribution after an infinite time,

$$P(\geqslant s) = s^{-\alpha} \tag{14}$$

which is a power law. The exponent $\alpha = [\ln(Z-1)]/\ln[(Z-1)\rho]$, which depends on the forward branching ratio $(Z-1)$ (i.e., number of offspring per parent cell) as well as the probability ρ that the offspring will be identical to the parent. For a given value of Z, the power law becomes steeper as ρ decreases from near unity towards the critical point. This reflects the fact that for smaller rates of mutation (larger ρ) large population sizes of a given strain are more likely. Whereas, when mutation rates are larger (ρ is smaller) the probability of a large population of a given strain is smaller. When $\rho \leqslant \rho_c$, this argument breaks down because the strains are no longer growing, and a power law is no longer observed.

Mathematically the mechanism that produces the power laws is associated with competing exponential growth and decay processes. Because the overall size of the population as a whole (all strains) grows exponentially with generation, the probability a given strain originates in generation t_b grows exponentially as well, i.e., $P(t_b)$ [Eq. (9)] grows exponentially in t_b. However, the number of offspring s in the current generation which results from a mutation in generation t_b, decreases exponentially with increasing t_b, reflecting the fact that smaller strain sizes are typically associated with more recent mutations. The competing exponential processes produce a power law in $P(\geqslant s)$ [Eq. (14)].

Exponentials of exponentially distributed variables comprise a general probabilistic mechanism leading to power law distributions. It has been discussed by Mandelbrot both generally and in the specific context of preferential growth [33]. In Sec. V we see that HOT generates power laws in a mathematically similar (but more general) way. We discuss both models in the context of this broader mathematical context in Sec. VII.

Next we return to the approximation made in Eq. (11) where we replaced the actual (random) growth of clone population size with deterministic growth at the average rate. Of course, the growth of clone populations is *not* really deterministic. Taken literally, Eq. (11) implies there is only one possible strain size s at the current generation t associated with mutations at generation t_b. In particular, since $(t-t_b)$ increases by unity each generation, Eq. (11) implies s increases by a factor of $(Z-1)\rho$ each generation. This discrete set of s values produces steps of width $\log_{10}[(Z-1)\rho]$ for the cumulative probability $P(\geqslant s)$ of strain sizes greater than or equal to s, plotted on a $\log_{10}$-$\log_{10}$ graph. Here the term "power law" describes the decay of the predicted values of sizes (the upper right corners of the steps). This is illustrated in the solid, piecewise constant curve in Fig. 3. It should also be noted that while an individual strain's size grows exponentially [grows by a factor of $(Z-1)\rho$ each time step, as in Eq. (10)], it grows more slowly than the number of boundary

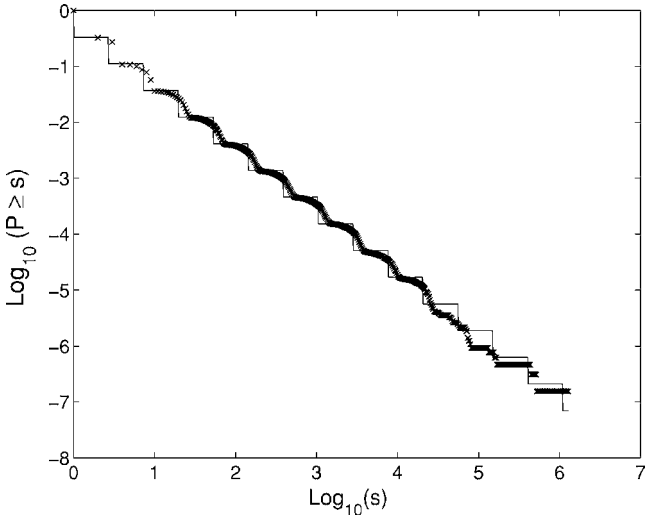

FIG. 3. Preferential growth. Cumulative number of strains $P(\geqslant s)$ vs strain size s with $Z=4$ and $\rho=0.9$ (well above the critical point). The solid lines are the deterministic analytic results, and the $\times$'s are the results of stochastic simulations. The predicted distribution has steps because there is only one strain size per generation (i.e., it is discrete because all strains created in a given level will have the same size). The numerical result is not exactly a step because fluctuations produce some variation in the sizes. In this case, we show results for $Z=4$ (instead of 3) to make the "steps" wider and easier to see.

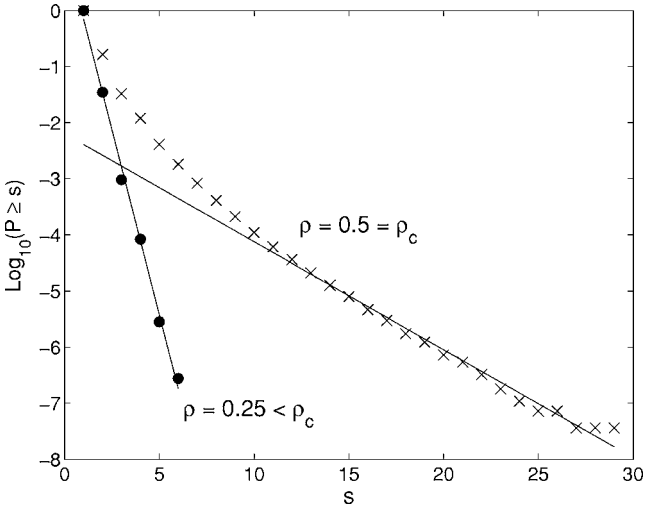

FIG. 4. Preferential growth. Cumulative number of strains $P(\geqslant s)$ vs strain size s with $Z=3$ and $\rho=0.5$ (at the critical point) for $\times$'s, and $\rho=0.25$ (below critical point) for the circles. Best fit lines are included in the log-linear plot to illustrate that the graphs are *not* power laws, and are in fact closer to exponential distributions. Note the difference in scale compared to Fig. 3.

sites of the whole lattice (which grows by a factor of $Z-1$ each time step). Thus in the $t\rightarrow\infty$ limit, any individual strain makes a negligible contribution to the total population (except in the case $\rho=1$ where of course the whole population is one strain, as there is no mutation).

When the assumption of deterministic growth used in the analytical calculation is replaced by the original random process [where mutation occurs randomly for each offspring with probability $(1-\rho)$], we observe a range of sizes centered around the average size given in Eq. (11). These stochastic effects smooth the steps. For large values of ρ, as illustrated in Fig. 3, the steps are still visible, and the analytical formula for $P(\geqslant s)$ [Eq. (13)] fits the data well. As ρ decreases toward ρ_c, the rounding in the steps increases, but the power law predicted in Eq. (13) remains a reasonably good fit.

At and below the critical point, the growth of the strains can no longer be approximated by the deterministic growth we described (below ρ_c the strain sizes tend to shrink rather than grow). Instead, the only way that a strain achieves a size greater than unity is through transient fluctuations. The average behavior of all strains is to decrease in size and eventually die. A reasonably good fit to the size distribution can be obtained by assuming that all the surviving strains were created in the last few generations, and calculating the probability that they would have achieved a certain size. There is no power law below the critical point. Instead the distribution has a roughly exponentially cutoff beyond some characteristic size. This characteristic strain size decreases with decreasing ρ. This can be seen numerically in Fig. 4.

IV. PERCOLATION

Percolation is among the simplest, most analytically tractable models in statistical physics that exhibits a critical phase transition [26]. In site percolation sites on the lattice are independently occupied with probability ρ and vacant with probability $(1-\rho)$. Thus ρ corresponds to the density of occupied sites. The critical density ρ_c is defined by the onset of long range connectivity of occupied sites. That is, ρ_c is the density marking the onset of nonzero probability that a chosen site in the interior of the lattice (the origin) is connected via a continuous path of occupied sites to the boundary, in the limit of infinite lattice size. The model can be solved exactly in $d=1$, and on the Bethe lattice, and many quantities are known exactly in the $d=2$ case. In this section we show the exact solution on the Bethe lattice, for completeness (and because it has been difficult to find some details of the calculations in the literature).

Criticality in percolation is closely related to the self-organized critical forest fire models [34,35]. Such models include rules that ultimately govern evolution of ρ through specified rates for occupying vacant sites (growing "trees"), introducing sparks which ignite trees, and propagating fires through connected clusters (which are then converted to vacancies). In the forest fire models, the lattice evolves to a critical density and the statistics of burned clusters are described by power laws. Here we discuss static percolation rather than the dynamic SOC models, and consider cluster size distributions as well as fire size distributions. In this case fire sizes are generated over the ensemble of static configurations rather than a dynamical algorithm.

Percolation configurations are generated using the same algorithm we used to generate preferential growth configurations in Sec. III. However, for percolation the interpretation of the lattice structure is spatial, rather than temporal, and we

are primarily interested in connected clusters in the interior of the lattice (the Bethe lattice, defined in Sec. II), rather than the boundary sites with a common ancestor on the Cayley tree.

In percolation, a connected cluster is defined to be a set of nearest neighbor connected occupied sites, surrounded completely by vacant sites at the perimeter. Strictly speaking, a given cluster is said to be "percolating" if it is infinite on an infinite Bethe lattice, as defined in Sec. II. On a large finite lattice, we will use the terminology "percolating clusters" to describe properties of a finite lattice as if it were embedded within an infinite lattice. We refer to any cluster on the finite lattice that comprises a subset of a percolating cluster on the infinite lattice as "percolating." In finite dimensions there is either one (above ρ_c) or zero (below ρ_c) percolating clusters on an infinite lattice. However, on a Bethe lattice, the absence of loops gives rise to many infinite clusters for densities above ρ_c.

In fact, boundary sites of these percolating clusters on a Cayley tree make up the populations of larger strains in preferential growth above the critical point. The large strains almost surely continue to grow without bound and therefore clearly result in an infinite cluster (on an infinite lattice). On the other hand, the probability that an infinite cluster will maintain a bounded strain size goes to zero as the lattice size goes to infinity.

The probability that a randomly chosen site is both occupied and part of an infinite cluster defines the percolation order parameter $P_\infty(\rho)$. This can be calculated on the Bethe lattice using a recursion relation for the probability that the origin is in a finite cluster $P_f(\rho)$, and noting that $P_\infty(\rho)=\rho - P_f(\rho)$. A given cluster is finite if and only if it has at least one occupied site, and all its sub-branches are finite:

$$P_f(\rho) = \rho \sum_{n=0}^{Z} \binom{Z}{n} \rho^n (1-\rho)^{Z-n} P_{fb}^n(\rho), \tag{15}$$

where $P_{fb}(\rho)$ is the probability that the sub-branch is finite. Here a sub-branch is defined to be a connected subcomponent of the original cluster, and thus share a common connection to the remainder of the cluster.

Computing the sum in Eq. (15) yields

$$P_f(\rho) = \rho(1 - \rho + \rho P_{fb})^Z. \tag{16}$$

Similarly, a sub-branch is finite if and only if its sub-branches are finite,

$$P_{fb}(\rho) = \sum_{n=0}^{Z-1} \binom{Z-1}{n} \rho^n (1-\rho)^{Z-1-n} P_{fb}^n(\rho), \tag{17}$$

which yields the implicit equation for $P_{fb}(\rho)$:

$$P_{fb}(\rho) = (1 - \rho + \rho \times P_{fb})^{Z-1}. \tag{18}$$

This equation can be solved numerically to obtain $P_{fb}(\rho)$ for an arbitrary given value of ρ. Notice that the right side of Eq. (16) is just ρ times the right side of Eq. (18) raised to the $Z/(Z-1)$ power. Using this, we obtain

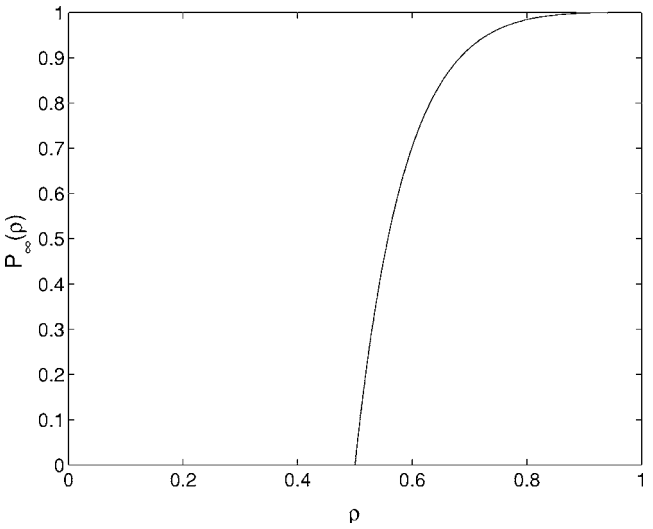

FIG. 5. Graph of the probability a cluster is infinite, P_∞ vs density (ρ) for $Z=3$. Note the jump in the first derivative at $\rho = 1/2$.

$$P_f(\rho) = \rho P_{fb}^{Z/(Z-1)}. \tag{19}$$

Since $P_\infty=\rho-P_f(\rho)$, we have

$$P_\infty = \rho(1 - P_{fb}^{(Z-1)/Z}). \tag{20}$$

Thus one can numerically solve Eq. (18) and substitute into Eq. (20) to find the value of the order parameter for all ρ. Below $\rho=\rho_c$, $P_\infty(\rho)$ is zero, but it has a jump in its first derivative at ρ_c, and monotonically increases after that point. Figure 5 shows an example plot of P_∞ vs ρ.

In statistical physics, a special significance of the critical density ρ_c is the emergence of power laws in the distribution of cluster sizes. In fact, for configurations generated randomly at fixed density, ρ_c is the only density which leads to power laws in the cluster size distribution in the interior of the lattice. The strain sizes are measured at the boundary of the Cayley tree, and occur over a range of densities (from ρ_c to unity), with a power law exponent α which decreases monotonically across this range. Together these calculations illustrate how power laws may arise in what is essentially the same model, but at different densities, in different quantities, and by different mechanisms. The goal of the remainder of this section is to calculate $P(\geqslant s)$, the probability that a randomly selected cluster has size greater than or equal to s, and compare the result with the distribution of strain sizes from the previous section on preferential growth.

On any lattice with filling probability ρ, the probability $\phi(s)$ that a given site is occupied and in a cluster of size s is defined to be:

$$\phi(s) = \sum_k s g_{sk} \rho^s (1-\rho)^k \tag{21}$$

where the sum runs over the cluster perimeter, indexed by k, consisting of unoccupied sites. Here g_{sk} is referred to as the "animal" number—the number of distinct clusters (up to translation) with fixed s and k (note that the extra factor of s is required because g_{sk} only describes the *shape* of the clus-

ter, and the given site could be in any of the s nodes of the cluster). On the Bethe lattice, a cluster with a given size always has the same number of sites composing the perimeter, $k=(Z-2)s+2$, so $g_{sk}=g_s$ and

$$\phi(s) = s g_s \rho^s (1-\rho)^{(Z-2)s+2}. \tag{22}$$

The noncumulative $P(s)$ (i.e., the probability that a cluster is of size s) can be related to the probability a given occupied site is in a cluster of size s as follows. The cluster density n_s (also called the "cluster number" [26]) is defined to be the number of clusters of size s per unit volume:

$$n_s = \phi(s)/s = g_s \rho^s (1-\rho)^{(Z-2)s+2}. \tag{23}$$

To see that this is correct, multiply $\phi(s)$ by V (the volume) to get the total number of sites in clusters of size s, then divide by s to get the total number of clusters of size s, then finally divide by V to get the cluster density. $P(s)$ is then simply the normalized density:

$$P(s) = n_s / \sum_{j=1}^{\infty} n_j. \tag{24}$$

What remains is to determine the animal numbers g_s in Eq. (21). The absence of loops on the Bethe lattice makes it possible to calculate g_s exactly. A detailed derivation of g_s is given in Appendix B. Here we simply state the result:

$$g_s = Z \frac{[(Z-1)s]!}{[(Z-2)s+2]!\,s!}. \tag{25}$$

This formula for g_s can be used to derive the cluster density n_s [Eq. (23)] and thus the cluster size density $P(s)$ [Eq. (24)]. Using the Stirling approximation, we obtain for large s

$$n_s \sim \frac{(1-\rho)^2}{\sqrt{2\pi}} \frac{Z\sqrt{Z-1}}{(Z-2)^{5/2}} s^{-5/2} e^{-cs} \tag{26}$$

where

$$c = -\ln \frac{(Z-1)^{Z-1}}{(Z-2)^{Z-2}} \rho (1-\rho)^{Z-2}. \tag{27}$$

And since $P(s) \propto n_s$, we can write $P(s)$ in the form

$$P(s) \sim s^{-\tau} e^{-cs} \tag{28}$$

with $\tau = \frac{5}{2}$. At $\rho = \rho_c \equiv 1/(Z-1)$, $c=0$, and if we Taylor expand c [Eq. (27)] for ρ near ρ_c, we find that as $\rho \to \rho_c$, c goes to zero as $c \sim (\rho - \rho_c)^2$. At ρ_c, $P(s)$ is a pure power law with exponent $-\tau$, which identifies ρ_c as the critical density where the average cluster size diverges, and defines the critical exponent $\tau = \frac{5}{2}$:

$$P(\geqslant s) = \sum_{j=s}^{\infty} P(s), \tag{29}$$

and since the sum of this discrete power law is well approximated by an integral, the cumulative distribution at the critical point is also a power law $P(\geqslant s) \sim s^{-\alpha}$, with an exponent of $\alpha = \tau - 1 = 3/2$.

Next we compare the power laws for percolation to those obtained in Sec. III for preferential growth. While preferential growth and percolation are both described by randomly filled or empty nodes at a fixed density, they produce different power laws, in different quantities, and at different densities. For percolation we measure the distribution of connected cluster sizes on the interior of the tree, and power laws occur with fixed exponent (independent of the coordination number), and extend across all length scales, with no cutoffs, only at the critical density. Approaching ρ_c from below the critical density, power laws exist over a limited range and retain the same exponent, $\alpha = 3/2$, up to a characteristic size, at which the size distribution cuts off sharply (exponentially). Above the critical density, there are power laws in the finite clusters, again with the same exponent, but again over a limited range, and percolating clusters also exist. Interestingly, unlike the finite dimensional case where the percolating cluster is unique when it exists, the absence of loops on the Bethe lattice leads to the existence of many percolating clusters. These percolating clusters, extend to the boundary sites on a Cayley tree, where the edge sites form the large strains in the present generation. The distribution of sizes of these strains is the quantity of greatest interest in preferential growth.

Figure 6 illustrates numerical results for the cumulative distributions of strain sizes and cluster sizes in preferential growth and percolation, respectively. Results for $Z=3$, a lattice with 26 levels, and a range of densities above and below the critical density $\rho_c = 1/2$ are shown. We plot cumulative results $P(\geqslant s)$, obtained by summing the probability density for sizes greater than or equal to size s. In (a) we show results for the strain sizes in preferential growth. At high densities (e.g., the results shown for $\rho = 0.85$) the power law is relatively shallow, compared to lower densities (e.g., $\rho = 0.55$), where the power law is relatively steep. As ρ decreases from near unity to ρ_c the slope of the corresponding data decreases smoothly. At ρ_c the power law ceases to exist, although we can still measure strain sizes. In this case, the strains are no longer associated with percolating clusters (there are none). Instead they originate in recent generations, and are associated with finite clusters which contact the lattice boundary in percolation. Their sizes, like the cluster sizes in percolation below ρ_c, cut off sharply beyond a characteristic size. Some strains also arise from finite clusters above ρ_c. However, compared to the strains associated with percolating clusters, they are small, and do not contribute significantly to the power laws. Of course on the Cayley tree there is no sharp distinction in the thermodynamic sense, because the lattice is finite. It is also of note that the power laws in preferential growth are the result of mean behavior, and may exist even if no randomness is present. By contrast, the power law distribution of event sizes in percolation is the result of stochastic deviation from mean behavior—if the process of filling the lattice was nonrandom, clusters formed would have a typical size, and the power law would be destroyed.

Figure 6(b) illustrates the analytical results for the distribution of finite cluster sizes in percolation for the same range of densities. Note the dramatic change in scale compared to Fig. 6(a). In this case, power laws are observed only for $\rho = \rho_c$. At lower densities, the same power law is observed for small cluster sizes, but the distribution cuts off sharply at a

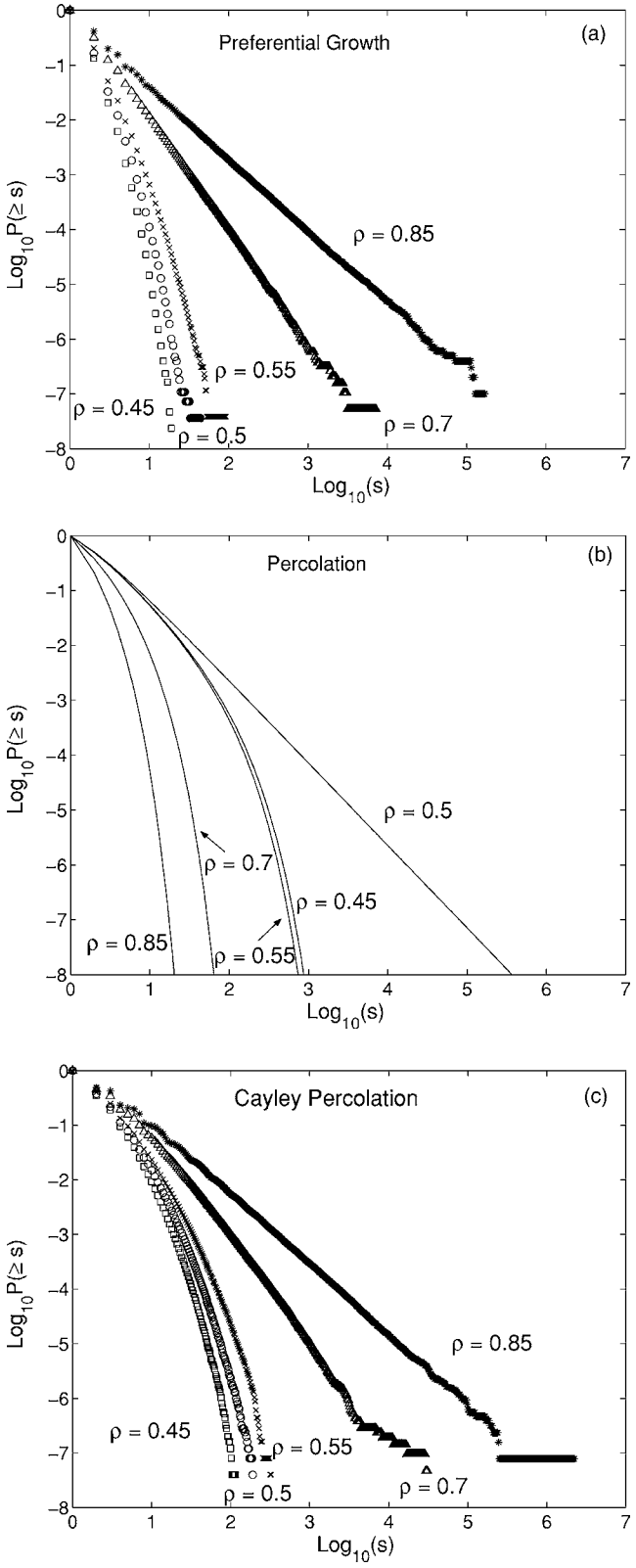

FIG. 6. Distribution of sizes for preferential growth (a), percolation on the Bethe lattice (b), and percolation on the Cayley tree (c) for a lattice with $Z=3$ and 26 levels. Squares are slightly below critical ($\rho=0.45$), circles are at critical ($\rho=0.5$), $\times$'s are slightly above critical ($\rho=0.55$), triangles are further above critical ($\rho=0.7$), and asterisks are even higher ($\rho=0.85$).

characteristic size which decreases with decreasing density. At high densities (e.g., $\rho=0.85$), the finite-cluster size distribution also cuts off at small sizes. In this case, a large portion of the occupied sites are associated with infinite clusters.

In Fig. 6(c) we illustrate results for percolation for all clusters on the Cayley tree (including both finite and percolating clusters without distinction). Thus (c) illustrates the cumulative distribution of sizes of all the clusters on the Cayley tree, regardless of their interaction with the boundary. At high densities (e.g., $\rho=0.85$), the percolating clusters dominate the statistics, and our results resemble those obtained for preferential growth. Note, however, that the largest cluster sizes are somewhat larger than the largest strain sizes, because the measure of size s in percolation includes the interior sites as well as the boundary. For densities near ρ_c the percolating clusters continue to dominate the tail of the distribution, and the distribution of sizes is a power law, with exponent given by the strain size distribution. As the density decreases below ρ_c the finite clusters dominate, and the size distribution is described by percolation.

On the Bethe lattice, percolation exhibits power laws near the critical point, where the percolating clusters are on the verge of forming and the system is arguably at its most complex; yet that is precisely where the power laws associated with preferential growth on the Cayley tree cease to exist. Another important distinction between the models is the overall size scale associated with the clusters and strains. In percolation, power laws are associated with finite clusters, only at the critical point does the power law extend to the size of the lattice, although even at criticality, the emergent percolating clusters are fractal, and scale sublinearly with the volume of the lattice (on the Bethe lattice, percolating clusters scale sublinearly with volume provided the density is less than 1). Note also that while the power law produced by percolation is fixed, the exponent in preferential growth is *tunable*. Other mechanisms (including HOT, as we will see) also have tunable exponents.

V. HIGHLY OPTIMIZED TOLERANCE

Highly optimized tolerance is a mechanism for complexity and power laws which incorporates a simplified notion of robust design. HOT is motivated by advanced technological systems and studied in that context using algorithms where design tradeoffs are deliberately optimized [13]. It is also motivated by biology and ecology, where algorithms based on natural selection lead to similar states in simple models [19,21].

HOT was initially introduced in the context of modified percolation forest fire models [14,15,20,36]. The simplest percolation forest fire model is obtained by adding the notion of "sparks" and "fires" to percolation. Occupied sites correspond to trees, and vacant sites correspond to fire breaks. When a spark lands on a vacant site, nothing happens. When a spark lands on an occupied site, a fire burns the tree on that site, as well as all sites in the corresponding nearest neighbor connected cluster.

These are the same clusters studied in the previous section. However, because there are more sites in the larger

clusters, they are more likely to be hit. As a result, the critical exponent for the power law describing the fire size distribution at criticality differs (by unity) compared to the exponent for the cluster size distribution. Specifically, on the Bethe lattice, where the probability $P(s)$ of a cluster of size s scales as $P(s) \sim s^{-\tau}$, with $\tau=5/2$ for $\rho=\rho_c$, and the cumulative distribution is $P(\geqslant s) \sim s^{-\tau+1}$ (see Sec. IV), then the probability $F(s)$ of a fire of size s scales as $F(s) \sim sP(s) \sim s^{-\tau+1}$, with corresponding cumulative distribution $F(\geqslant s) \sim s^{-\tau+2} \sim s^{-\alpha}$. This defines the exponent α for fires. On the Bethe lattice, this results in the exponent $\alpha=1/2$ for the cumulative distribution of fires of size greater than or equal to size s at the critical point $\rho=\rho_c$.

Note that for this simple percolation forest fire model, the result is largely independent of the spatial distribution of sparks—they could always fall on the same lattice site, be completely random, or anything in between, and it would not change the result (one exception to this would be changing the distribution to be very near the lattice edge versus near the middle, as clusters near the edge tend to be somewhat truncated). This is because the random ensemble of configurations at any given density ρ is translationally invariant. In the HOT version of this model, the detailed configuration on the lattice is optimized (subject to constraints) to maximize the number of trees remaining after a single fire, based on a given distribution of sparks. This results in special, *robust* configurations, which on average retain many more trees than their random counterparts. However, in HOT the lattice layout is also sensitive to the assumptions about (or history of) sparks. If a lattice is designed for a particular distribution of sparks, and the distribution of sparks changes or there were errors in the assumed distribution, then the performance of the designed lattice may deteriorate significantly. This is a *fragility* which is not present in the random case. This *robust, yet fragile* behavior is a key signature of the HOT mechanism. We compare the results for HOT configurations with the corresponding percolation forest fire model in Sec. VI.

To optimize the HOT lattice, we return to the Cayley tree for two reasons. First, working on a finite lattice allows us to work with finite quantities (e.g., cluster size s is always finite), making both analytic calculations conceptually simpler, and numerical computation feasible with few approximations. Second, as we will see, loss grows sublinearly in lattice volume (this is true for percolation and HOT on the Cayley tree) which introduces subtleties in the thermodynamic limit. In particular, in the limit of infinite lattice sizes, both percolation and HOT have no macroscopic loss. Nonetheless, many of the essential differences remain well illustrated on the Cayley tree.

To set up the optimization problem, we define a function which reflects the performance or fitness of a configuration. Typically, this is the fractional yield $\mathcal{Y}$ which is the average density of trees remaining after a single spark hits the lattice and burns through the corresponding connected cluster. The average loss associated with a given configuration is computed over a prespecified distribution of sparks. For the initial calculations performed in this section, we take the spark distribution to be peaked at the center of the lattice, decaying exponentially approaching the edge. This choice has the advantage of making the calculations analytically tractable, and illustrating the connection between the power laws which arise from the HOT mechanism and the functional "exponentials of exponentials" mechanism which arose previously in the context of preferential growth. After that we will consider other spark distributions, which also produce power laws, and suggest a generalization of exponentials of exponentials mechanism to other functional forms.

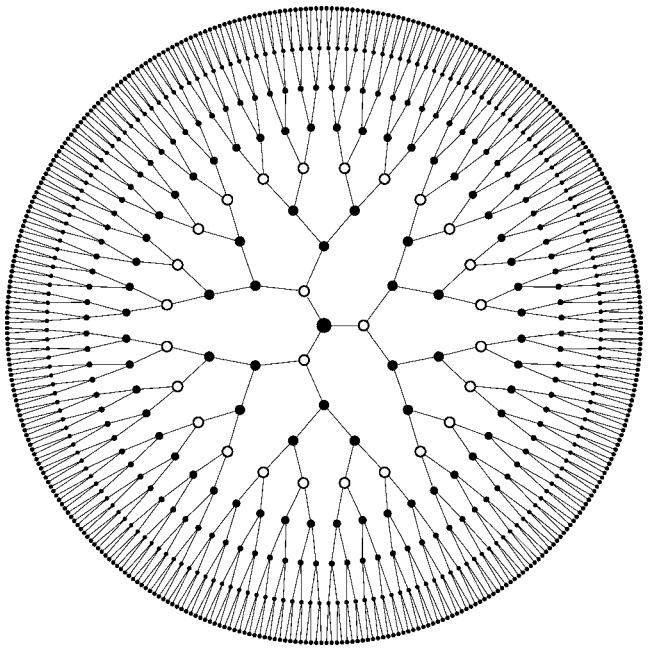

FIG. 7. A typical HOT configuration for a tree with $Z=3$. The empty circles are barriers.

We assume the probability that the spark lands on a node in the nth level is given by $p(n) \sim e^{-\kappa n}$. (This choice makes analytic solutions tractable.) Because of the radial symmetry of the spark distribution, the lattice layout which optimizes yield (defining the HOT configuration) consists of a series of concentric rings of vacant nodes at particular distances from the origin (the center node), on an otherwise fully occupied lattice. Each pair of consecutive fire break rings (vacancies) defines a family of connected clusters, all of equal size, between the inner and outer ring. Moving from the origin outward, the cluster sizes become progressively larger, as the probability of a spark becomes less likely. Figure 7 illustrates a sample configuration with rings at $n=1$ and 4. What remains is to calculate the specific positions of the concentric fire break rings, given a particular $p(n)$.

Suppose the Cayley tree has $N+1$ levels (labeled 0 to N), where N is large. Suppose M of the levels are empty (call the empty levels $\{c_i\}$) and fill the rest. Then the net yield Y can be written in terms of the $\{c_i\}$ as

$$\mathrm{Y} = 1 + Z\frac{[(Z-1)^N - 1]}{(Z-2)} - \sum_{i=1}^{M} Z(Z-1)^{c_i} - \langle L \rangle, \qquad (30)$$

where the first two terms count the total number of sites in the Cayley tree, the sum over i subtracts the vacant sites, and the average loss $\langle L \rangle$ remains to be computed. Note: here we

TABLE I. Probabilities and losses from sparks landing in various annuli of a Cayley tree with M cuts. The regions are labeled by n from 0 to M (i.e., s_2 is the loss in the third region) with regions 0 and M listed separately because their formulas are slightly different. For this table, $\psi(n) \sim \exp(-\kappa n)$.

Loss	Probability
$s_0 = 1 + Z\dfrac{(Z-1)^{c_1-1}-1}{Z-2}$	$P_0 = \dfrac{1-e^{-c_1\kappa}}{1-e^{-(N+1)\kappa}}$
$s_n = \dfrac{(Z-1)^{c_{n+1}-c_n-1}-1}{Z-2}$	$P_n = e^{-c_n\kappa}\dfrac{1-e^{-(c_{n+1}-c_n-1)\kappa}}{1-e^{-(N+1)\kappa}}$
$s_M = \dfrac{(Z-1)^{N-c_M}-1}{Z-2}$	$P_M = e^{-(c_M+1)\kappa}\dfrac{1-e^{-(N-c_M)\kappa}}{1-e^{-(N+1)\kappa}}$

measure Y in terms of the total number of trees after a single spark, not the tree density $\mathcal{Y}=Y/V$, i.e., we have not yet divided by the number of sites on the lattice.

To compute the expected loss $\langle L\rangle$, define $P_i=\sum_{n=c_i+1}^{c_{i+1}-1}\psi(n)$ to be the probability of a spark landing between c_i and c_{i+1}, and s_i to be the size of the clusters between c_i and c_{i+1}. Then s_i is the loss if a spark lands anywhere in that region, and maximizing yield is the same as minimizing the cost function

$$J = \sum_{i=0}^{M} P_i s_i + \sum_{i=1}^{M} Z(Z-1)^{c_i}. \tag{31}$$

Table I summarizes the values of s_i and P_i for the exponential spark distribution, in terms of the cut positions (which remain to be determined).

Next we determine the locations of the fire breaks, which we can then use to determine the relationship between event probabilities and event sizes. We set $\partial J/\partial c_i=0$, and take c_i to be large. After a bit of algebra, the result is

$$c_i \sim \left(2 + \frac{\kappa}{\ln(Z-1)}\right)^i. \tag{32}$$

This allows us to relate c_{i+1} to c_i to obtain a simple expression for s_i and P_i,

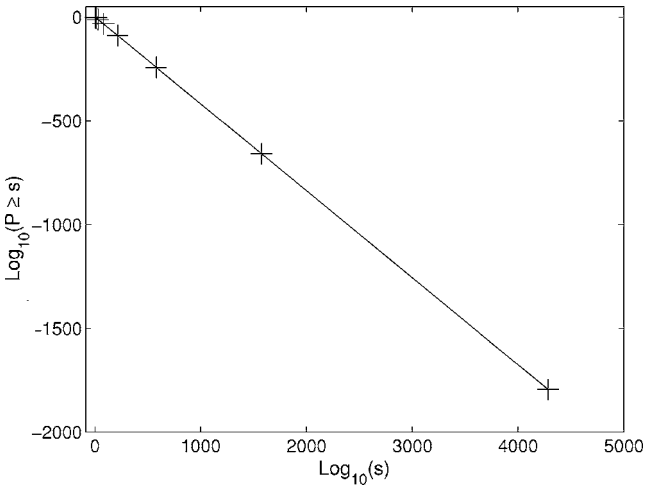

FIG. 8. HOT power law for $Z=3$ and $\kappa=0.5$, with a 22 500-level lattice. In this case the distribution is discrete. We show a line connecting the optimized event sizes (rather than steps) to illustrate the power law decay.

$$\ln s_i \sim [\kappa + \ln(Z-1)]c_i, \tag{33}$$

and

$$\ln P_i \sim -\kappa c_i. \tag{34}$$

Therefore

$$\ln P_i \sim \left(\frac{-\kappa}{\kappa + \ln(Z-1)}\right)\ln s_i \tag{35}$$

and

$$P_i \sim s_i^{-\kappa/[\kappa+\ln(Z-1)]}. \tag{36}$$

Figure 8 shows a plot of event sizes produced numerically with the calculated power law overlaid.

As in other cases, we are interested in the cumulative distribution $P(\geqslant s)$. In this case, the optimized values of s, $\{s_i\}$ grow so quickly that the sum rapidly becomes trivial:

$$P_{i+1} = \exp(-\kappa c_{i+1}), \tag{37}$$

and using Eq. (32),

$$P_{i+1} = \exp\left[-\kappa c_i\left(2 + \frac{\kappa}{\ln(Z-1)}\right)\right], \tag{38}$$

and thus

$$P_{i+1} = P_i^{2+\kappa/\ln(Z-1)}. \tag{39}$$

Successive P_i values decrease so rapidly that for large i, $P_i \sim \Sigma_{j\geqslant i}P_j$. Thus, the cumulative distribution is asymptotically the same as the noncumulative density

$$P(\geqslant s) \sim P(s_i), \tag{40}$$

where i is the largest integer so that $s_i \leqslant s$.

These calculations show that power laws in the HOT mechanism can arise from the functional mechanism involving exponentials of an exponentially distributed random variable which arose previously in preferential growth. The like-

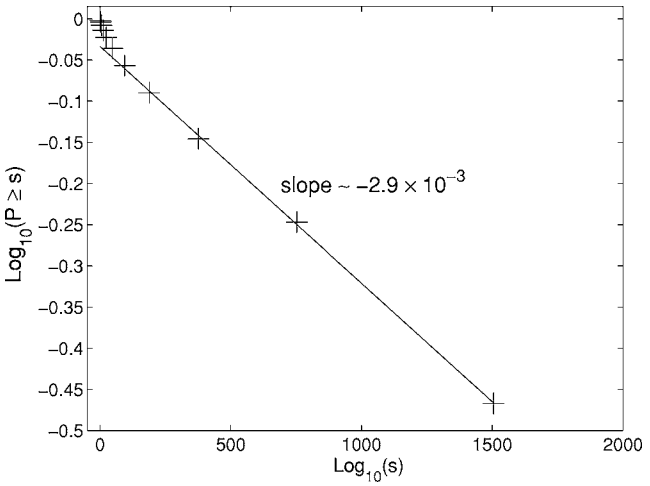

FIG. 9. HOT power law produced by a spark distribution with $\psi(n) \sim n^{-0.4}$ on lattice with $Z=3$.

lihood of a spark landing decreases exponentially with level, but the breaks are chosen by the optimization so that loss (a function of the randomly chosen level n) increases exponentially with level. Here again we have competing exponential growth and decay processes. Loss is the exponential of an exponentially distributed random variable; and thus has a power law distribution.

Furthermore, for HOT this basic functional mechanism appears to extend beyond the specific case of exponentials of exponentials. Indeed, a wide variety of spark distributions (not just exponentials) on a wide variety of lattices (not just the Cayley tree with its exponentially expanding boundary) lead to power law statistics for the HOT configuration [37]. For example, Fig. 9 shows a power law produced by HOT with a nonexponential spark distribution. In this case, the probability of sparks landing on the nth level, $\psi(n)$ was itself chosen to have a power law distribution:

$$\psi(n) \sim n^{-0.4}. \tag{41}$$

In this case the distribution rounds up somewhat at smaller sizes. The exponent α describing the tail is roughly $\alpha \approx 3 \times 10^{-3}$. The reason for such a shallow slope is twofold. First, power laws decay more slowly than exponentials, so more events take place farther from the origin of the lattice. Second, a very flat power law was chosen for the spark distribution so as to increase the number of barriers and better illustrate the distribution of events; however this also increases the number of events far from the origin. Having many events far from the origin makes containing them more difficult, and thus increases the probability of large events. More reasonable (steeper) spark distributions also result in steeper power law event distributions but give few data points with the computational resources we possess. Thus such spark distributions are somewhat less useful as examples for exhibiting the asymptotic power law in the distribution of event sizes.

The presence of a power law for a decaying but nonexponential distribution of sparks suggests a possible generalization of the function exponentials of exponentials mechanism. Let X be a random variable with cumulative probability distribution

$$P(X \geqslant x) = F(x), \tag{42}$$

and let the random variable Y be a monotonically increasing function of X

$$Y = H(X). \tag{43}$$

Formally, we can determine the distribution for Y:

$$P(Y \geqslant y) = P(H(X) \geqslant y) = P(X \geqslant H^{-1}(y)), \tag{44}$$

so that

$$P(Y \geqslant y) = F(H^{-1}(y)). \tag{45}$$

In HOT, $F(x)$ is the specified spark distribution and $H(X)$ results from optimization. Numerical results on the Cayley tree [and on $(d=1,2)$-dimensional lattices and with a variety of choices of ψ] suggest that through optimization, HOT picks an $H(X)$ that will produce a power law for $P(Y \geqslant y)$. Of course, this is not *always* possible. Two simple examples where HOT cannot produce power laws are the following.

(1) $F(x)$ constrained to a region of the lattice near the origin (a δ function or step function). Clearly there is no $H(X)$ so that $F(H^{-1}(y))$ is a power law. When fires are nearly always ignited from a particular site or within a constrained region, optimal solutions simply isolate the region or fill it with vacancies. HOT optimization then essentially focuses resources on only one ignition site or local region, and hence no power law will be observed.

(2) $F(x) \sim 1 - x/x_{max}$, which arises in the case of a uniform probability density (i.e., all ignition sites equally likely). Again, there is no $H(X)$ that can create a power law. Although there may be many different possible events, they are all equally likely, and hence optimization will assign each exactly the same loss. Thus there is only one event size.

However, our numerical results for exponential, power law, and other spark distributions, which peak at the origin, and decay approaching the boundary of the Cayley tree, suggest that optimization for yield Y often leads to power laws in the distribution of events. While this is not a proof, it suggests that the HOT optimization process may pick a particular H for a given F in a manner which leads to power laws. In the exponential case described above, $F(x) \sim e^{-\kappa x}$ and $H(x) \sim [\ln(Z-1)+\kappa]x$ producing the power law for $P(Y \geqslant y)$, which is the familiar exponentials of exponentials mechanism. We expect the more general mechanism in Eq. (45) occurs when $F(x)$ is not an exponential.

VI. COMPARISON OF HOT AND RANDOMLY FILLED LATTICES

How do the specialized HOT configurations compare with the randomly generated configurations at fixed density which describe percolation and preferential growth? Clearly HOT configurations are highly specialized, with a high density of vacancies (fire breaks) in regions where sparks are common, and low densities in regions where sparks are rare. Compared to the other models, where all configurations at a given den-

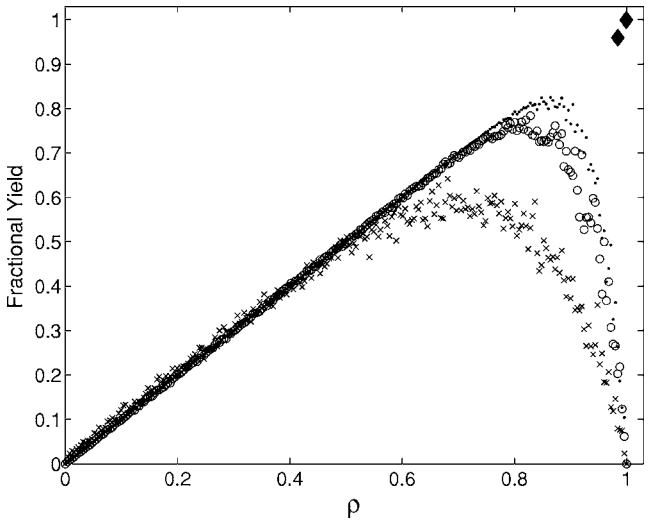

FIG. 10. Graph of fractional yield vs density for percolation on lattices with $Z=3$ and $\kappa=1$. The percolation data are for lattices with six ($\times$'s), 14 (circles), and 20 levels (dots). HOT data are represented with diamonds. The 14- and 20-level HOT lattices are both essentially at $\mathcal{Y}=\rho=1$, the six-level HOT lattice is at slightly lower $\mathcal{Y}$ and ρ. The critical point is at a density of $\rho=1/2$. Note that the axes extend slightly beyond $\mathcal{Y}=1$ and $\rho=1$ for clarity.

sity are *a priori* equally likely, HOT configurations are rare and specialized, and extremely sensitive to changes in the distribution of sparks.

In order to make a quantitative comparison of the performance of the models, it is useful to consider random lattices subject to sparks, and compute the yield in a manner analogous to HOT. This defines a percolation forest fire model, where yield can be calculated as a function of density. For a given lattice and a particular spark, fractional yield $\mathcal{Y}$ is the fraction of occupied sites which remain after the spark lands on the lattice and burns through the corresponding connected cluster. The average yield is computed by averaging over the distribution of sparks and the ensemble of random lattices at a given density. The results are shown in Fig. 10.

For the random lattices at low densities, most of the clusters are small (the average cluster size scales sublinearly with the system size), and the loss is insignificant. Thus fractional yield is essentially equal to the average density in this regime. Percolating clusters which span the system size first emerge at the percolation critical point. For finite-dimensional lattices, this also defines the maximum yield point for random configurations (this is not the case for the Bethe lattice), and yield is simply related to the percolation order parameter $P_\infty(\rho)$ (Fig. 5), which is the probability a given site is in the percolating cluster (which is unique):

$$\mathcal{Y}(\rho) = [1 - P_\infty(\rho)]\rho + P_\infty(\rho)[\rho - P_\infty(\rho)]. \tag{46}$$

At the critical density (in finite dimensions) the percolating cluster is a fractal. When a spark lands on the percolating cluster it results in a fire which spans the system, but with no net loss in tree density in the forest. Above the critical density, the percolating cluster comprises a finite fraction of the total density.

This is not the case for the Bethe lattice. Unlike finite-dimensional lattices, the Bethe lattice has no closed loops. Above the critical density for percolation, this results in not one (as in finite dimensions) but many percolating clusters. None of these clusters contains a finite fraction of the mass of the lattice.

To see this, consider a Cayley tree with n levels, and increase n in steps. As we saw in Sec. III on preferential growth, on the Cayley tree above the critical point, typical strains grow like

$$s(t) \sim [\rho(Z-1)]^n. \tag{47}$$

Strains are the boundaries of percolating clusters, hence on average clusters grow at the same rate. However, the number of nodes in the whole lattice grows like

$$N(t) \sim (Z-1)^n. \tag{48}$$

Thus, the fraction of the lattice occupied by any particular percolating cluster decays as

$$\frac{s(n)}{N(t)} \sim \frac{[\rho(Z-1)]^n}{(Z-1)^n} = \rho^n. \tag{49}$$

Note that the fraction of sites that are in percolating clusters remains constant [see Eq. (20)] but the fraction of sites that are in one particular percolating cluster approaches zero as n goes to infinity. Percolating clusters on the Bethe lattice and Cayley tree do not inevitably merge, as they do in finite dimensions, because of the lack of loops.

This is reflected in the fact that the fractional yield for random lattices in Fig. 10 does not show a significant drop off or decrease at the critical density $\rho_c=1/2$. On a finite-dimensional lattice, above the critical point there is a nonzero chance of macroscopic loss when a spark hits the percolating cluster. On the Bethe lattice, even if a percolating cluster is hit, the resulting loss is infinitesimal in the limit of infinite lattice sizes. Thus the observed drop off in yield in Fig. 10 is a consequence of the finite size of the lattice, and in the limit of infinite lattice sizes, the maximum fractional yield approaches the limiting value of unity.

Nonetheless, for finite-sized lattices, at very high densities the yield does drop off. The yields from percolation and HOT are compared in Fig. 10. For HOT, the fractional yield is essentially unity (note the slight offset in the boundary of the figure—results for lattice sizes 14 and 20 are both essentially at $\mathcal{Y}=\rho=1$), which is the same as in the finite-dimensional case. In both cases, high density configurations are achieved, with minimal losses due to optimal placement of fire breaks. On finite-dimensional lattices, the fire breaks consist of $(d-1)$-dimensional perimeters surrounding compact, contiguous, modular clusters that scale with the d-dimensional volume.

How does HOT achieve optimality? The HOT system places its largest clusters near the edge—where there are the most clusters, and where they are least likely to be sparked. Thus the typical cluster size is nearly as large as the maximum cluster size, yet the typical event size is much smaller. The sparks predominantly land near the center of the lattice,

TABLE II. Characteristics of the three mechanisms: preferential growth, criticality, and HOT.

	Preferential growth	Criticality	HOT
Configuration	Random	Random	Optimized
Connectivity	Time	Space	Space
Correlations	Random	Scale-free	Modular
Power laws	Strains	Clusters	Fires
At densities	$\rho_c < \rho < 1$	ρ_c	$\rho \approx 1$
Exponent	Variable	Fixed	Variable
Sensitivity	Low	Low	High
Functional origin	Deterministic or stochastic (growth vs decay)	Intrinsically stochastic (fluctuations)	Deterministic or stochastic (optimization)

on smaller clusters. Relatively few nodes are empty, so the system is close to a density of one and a fractional yield of one.

The HOT optimality comes with a price: potentially extreme sensitivity to the distribution of sparks. The yield from percolation will not change substantially if the parameter κ changes, or even if the general shape of the distribution $\psi(n)$ is changed (say, to make all the sparks land near the outside edge). However, the spacing of the barriers in HOT depends sensitively on the parameters of the spark distribution. If the distribution was changed to make the sparks land near the edge, the HOT scheme would be completely foiled. This is a common theme with HOT systems: robustness to designed-for uncertainty increases fragility to rare events, and changes in the disturbance pattern.

VII. CONCLUSIONS

In this paper, we examine three mechanisms for power laws in parallel using one lattice model: a Cayley tree with nodes that are filled or empty. The Cayley tree allows us to consider the models in a unified format, where concepts and calculations are (relatively) straightforward. Additionally, presenting the mechanisms in parallel emphasizes the intrinsic similarities and differences. In this paper, we primarily show plots for coordination number $Z=3$ to aid in comparisons; however, most of the derived equations generalize to arbitrary Z. HOT arises when the lattice is deliberately optimized for high yields, which leads to a specific, nongeneric, modular lattice layout, which has high performance, but is also sensitive to any change in the details of the optimization problem which it is designed to solve (e.g., the particular spark distribution). Percolation and preferential growth are both defined using random lattices, so their properties and statistics are representative of generic ensembles of lattices at a given density. As a result they are much less sensitive to changes in model details, but they are also far from optimal in the context of almost any optimization problem. The difference between percolation and preferential growth lies not in the specific configurations which are studied, but rather in the quantities examined and the interpretation of the lattice geometry in terms of space vs time. A summary of these and other distinguishing characteristics is given in Table II.

Of course, there are many purely statistical mechanisms that robustly produce power laws (see Appendix A). The discovery in data of high variability or scaling by itself is not *a priori* suggestive of or a signature for any particular mechanism. As a result, it is important to understand the full range of mechanisms that may be responsible for high versus low variability in real data, or the presence or absence of length scales in real processes.

Some mechanisms predict a particular exponent α for the power law $P(\geqslant x) \sim x^{-\alpha}$. For example, return times of random walks are power laws with $\alpha=1/2$. Another example, $P(Y \geqslant y)$ for $Y=1/X$ and nontrivial support for $P(X<\epsilon)$ for small ϵ, gives $\alpha=1$. In the cases we have considered, criticality leads to a specific, fixed exponent (which in general depends only on long wavelength features of the system, such as lattice dimensionality, dimension of the order parameter, and range of interaction), while preferential growth and HOT have variable exponents which can be *tuned* by varying the density or branching ratio (preferential growth) or the spark distribution (HOT).

By far the simplest and most versatile mechanism for generating arbitrary α is the "exponentials of exponentials" we have discussed in the context of preferential growth and HOT. Written abstractly as a functional mechanism, it simply states that if $Y=e^{aX}$ and $P(X \geqslant x)=e^{-bx}$, then $P(Y \geqslant y)=P(X \geqslant \ln(y)/a)=y^{(-b/a)}$. This need not involve probability distributions but can be purely deterministic. For example, if we have two exponential functions of a common variable, such as $x(t)=e^{at}$ and $y(t)=e^{-bt}$, then eliminating t gives a power law $x=y^{-\alpha}$ with $\alpha=a/b$. Given that exponential functions and distributions are readily generated with diverse models, such as solutions of linear differential equations, it is easy for these elements to combine and produce a large variety of models with power law statistics. Add to this the strong statistical invariance properties of power laws, it is then possible to incorporate stochastic elements that preserve power laws.

One long standing and frequently used "exponentials of exponentials" model is the preferential growth model, a variant of which we have solved here. As presented by Mandelbrot [33], one early example of research in this area was the work of Yule [22], who in 1925 developed power law models to explain the observed distribution of species within plant genera. Luria and Delbrück in 1943 developed a model and

supporting mathematics for the explicit generation of scaling relationships in the number of mutants in old bacterial populations [7]. A more general model of preferential attachment was developed by Simon [38] in 1955 to explain the observed presence of power laws within a variety of fields, including economics (income distributions, city populations), linguistics (word frequencies), and biology (distribution of mutants in bacterial cultures). Substantial controversy and attention surrounded these models in the 1950s and 1960s [33]. A recent review of this history can also be found in [39].

Note that in preferential growth on the Cayley tree, the value of the exponent α in the power law is determined by the coordination number Z of the tree, and the probability ρ that the offspring is identical to the parent. The power law arises from the competition between the exponential production of new strains, and the exponential growth of existing strains. Preferential growth can be defined in the context of this exponentials of exponentials mechanism without specific reference to the Cayley tree as follows.

Consider a cell population, in which the number of clones in each strain is growing at a rate which depends on the current population size, and new strains are created during this growth process through mutation. The growth process is defined by the following rules.

(1) An index i defines each strain, with $1 \leqslant i \leqslant N$, and N defines the total number of strains at a given time.

(2) There is an ordered list of strain sizes $\{s_i\}$.

(3) At time $t=0$, there is a single cell, indexed as $i=1$ with $s_1=1$.

(4) In each subsequent time step, all existing strain sizes grow, $s_i \rightarrow s_i \times G$, with G an integer.

(5) At the same time as the growth occurs, there is an increase in the number of strains (representing mutation): $N \rightarrow N \times M$, with M an integer. All *new* strains are initiated with population size equal to 1, i.e., $s_i=1$ for $N < i \leqslant MN$.

As with preferential growth on the Cayley tree, this generates a process which can be described by exponentials of exponentially distributed variables. The probability that a randomly selected strain has a size at least s can be derived as follows:

$$P(\geqslant s) = P(G^{t-t_b} \geqslant s), \tag{50}$$

where t_b is the time at which the strain was first created (through mutation). Equivalently,

$$P(\geqslant s) = P\left(t_b \leqslant t - \frac{\ln s}{\ln G}\right). \tag{51}$$

The number created at time t is M^t, so

$$P(\geqslant s) = \frac{M^{t-\ln s/\ln G}}{M^t} \tag{52}$$

so that

$$P(\geqslant s) = M^{-\ln S/\ln G} = s^{-\ln M/\ln G}. \tag{53}$$

Since M and G are arbitrary integers, the resultant power law has an exponent that can be tuned to any rational number. (To see this, consider $G=2^p$ and $M=2^q$ with p and q integers.)

This abstract example shows how power law statistics can arise simply from having an exponentially growing number of objects, each of which is growing exponentially in size. No further details, such as geometry, graphs, or specially tuned rates, are necessary. These features may constrain the specific exponent in the power law. If strains grow like $s \sim G^t$ and the number of strains grows like $N \sim M^t$ then the strain size distribution will be a power law with exponent $\alpha = -\ln M/\ln G$. This deterministic growth process also shows that random fluctuations are not a necessary feature to obtain power laws. Furthermore, even completely discrete growth can result in arbitrary power laws (obviously smooth, continuous growth can as well).

For researchers unfamiliar with these mathematical, statistical, and data analytic issues regarding high variability in data, the ubiquity of power laws may be unexpected. Indeed, the discovery of such properties of complex systems and the ability to describe them with power law type relationships or scaling distributions has been a central theme underlying the attempts by researchers to understand and explain complexity [5,11,40]. One result is that in the 1990s, much of the science literature on complex systems focused on models based on critical phenomena from statistical physics [5].

Systems with phase transitions have a critical point where some average length scale goes to infinity, giving rise to scale free statistics and power laws. At other densities, the statistics exhibit an exponential cut off (independent of system size). However, at the critical point, the system looks statistically identical on all scales. For instance in percolation, if one were to look at one section of the lattice, then "zoom out" and view a larger portion, the two would be (statistically) indistinguishable at the critical point. This is because at the critical point, the system has no average size for clusters. Thus, the probability distribution of clusters must allow for clusters of arbitrarily large size, and it must have an infinite average size—in short, a diverging correlation length and a scale-free power law. This is a generic property of critical systems; a length scale that goes to infinity at a critical point, which gives rise to scale-free power laws. An infinite event size implies infinite variance as well, thus fluctuations in event sizes are also large. In some systems, these fluctuations can be observed directly, for example, through the phenomenon of *critical opalescence*. Note that unlike the other two mechanisms we consider, power laws at criticality are specifically associated with fluctuations and randomness. They cannot be approximated using a deterministic, functional mechanism based on the average density of occupied sites on the lattice, which would predict clusters all of equal size.

Critical phenomena only exhibit power laws over a narrow range of phase space near the critical point; but if the length scale or *order parameter* is not free, but governed by the evolution of the system, the power laws may be broadly observable—this is called self-organized criticality [5]. A simple example is the SOC forest fire model. "Trees" are periodically randomly placed on a lattice, and "sparks" strike the lattice randomly. If a spark strikes a tree, that tree and the cluster of trees connected to it "burn" and are removed from

the lattice. This is like percolation but with dynamics that control the density of filled sites. In a typical SOC model (e.g., on a finite-dimensional lattice), if the density were above the critical point, loss would be a finite fraction of the lattice size. So if the density were above critical, it would (on average) decrease. Below the critical point there are no large clusters that scale with the system volume, so loss would be infinitesimal, and thus the density would (on average) increase. Thus the system tends toward the critical density, where the distribution of cluster sizes is a scale-free power law. This leads the distribution of fires to be a power law as well.

Preferential growth as a mechanism for power laws resurfaced recently in the scientific literature in the context of scale free networks [12]. Modifications to the original Barabási-Albert construction have also been proposed and have resulted in scale free network models that reproduce power law degree distributions with any $\alpha \in [1,2]$, a feature that is claimed to agree empirically with many observed networks [27]. In this paper, we focus on the original models of preferential growth in branching processes such as bacterial populations using a simple model of percolation on a Cayley tree. Here the population is "grown" by simple branching with some small probability of mutation into a new strain, assumed to have the same growth rate. Then the strain sizes are the boundary sizes at large radii of the clusters in percolation. For all sufficiently small mutation rates, this distribution is scaling as the radius goes to infinity. This appears to be a more robust mechanism for generating power laws than criticality in the sense that it applies to a broad range of lattice densities. Note that in this case, there is no special role played by a diverging correlation length. While the largest strains naturally encompass a broader span (e.g., left to right) on the boundary, there is no special interpretation to (left vs right) spatial position for preferential growth. The lattice can be left out of the problem entirely [as in Eqs. (50)–(53) above, which make no reference to the relative position of the cells].

The final mechanism we have considered that produces power laws is HOT. A HOT system is subject to a variety of uncertain stresses, and attempts to optimally respond to them with limited resources. The tradeoffs from prioritizing allocation of resources typically result in power law statistics. It should be emphasized that these statistics do *not* result from an absence of scale in HOT. HOT systems are *scale rich*. "Zooming out" on a HOT system typically results in a very different appearance, revealing structures which were not visible on smaller scales. Correlations within the system are characterized by modular structures, isolating or compartmentalizing the system into separate regions using efficient barriers, which protect the system by limiting losses. These barriers may either be physical barriers like the fire breaks in the HOT forest fire model, or dynamical barriers (e.g., switches) in the state space of a system [19].

Optimality of HOT systems can sometimes be deduced by directly observing the process or algorithm which selects the system as a best solution to some biological or technological design problem. However, in complicated systems it is often difficult to determine exactly what is being optimized or why. One result of optimization should be observable: order. However, unlike the traditional order associated with phase transitions (e.g., percolation) in statistical physics, order in HOT is perhaps more appropriately referred to as organization, and corresponds to typically heterogeneously structured, extreme low entropy states, comprising to a set of measure zero within the ensemble of all possible configurations at a given density.

A HOT system is optimal, and therefore the best (or nearly the best) at solving some problem. Randomly rearranging a HOT configuration results in dramatic drops in performance. This is in sharp contrast with criticality or preferential growth configurations which are randomly generated. Randomly rearranging a percolation lattice results in *no change* in its statistics. The only time that a percolating lattice could be said to be ordered in the sense that HOT configurations are ordered is in the trivial cases where it is completely full or completely empty.

This order or organization is always defined in relationship to some external condition. A HOT configuration is optimal for a system coupled to an uncertain, perturbing environment (here a specific spark distribution). Tradeoffs are made in a manner that maximizes robustness for the spectrum of uncertainty, whether the uncertainty is directly specified, estimated, or experienced on evolutionary time scales. As a result, the performance of a HOT system deteriorates dramatically if the problem (e.g., rules or external conditions) are changed. In this way, HOT systems are extremely sensitive, or fragile to changes or uncertainties in model assumptions. For example, the HOT forest fire model established concentric rings around the most likely spark locations. If the distribution of sparks were to change, the yield for that design would degrade considerably. In contrast, a randomly filled lattice would be unaffected by such a change.

Often it is not immediately obvious which of these mechanisms, if any, are responsible for an observed power law. Both SOC and HOT have been proposed as explanations of statistics in forest fires; and all have been used to model web and Internet traffic. Both HOT and preferential growth invoke some sort of design or evolution, which can be represented or approximated as deterministic processes, but also allow for (somewhat) random growth. Only careful analysis can reveal which aspects are driving observed statistics.

ACKNOWLEDGMENTS

This work was supported by the David and Lucile Packard Foundation, NSF Grant No. DMR-9813752, the James S. McDonnell Foundation, and the Institute for Collaborative Biotechnologies through Grant No. DAAD19-03-D-0004 from the U.S. Army Research Office.

APPENDIX A: STATISTICAL PROPERTIES OF POWER LAWS

In this appendix we provide some background on the statistics of power law distributions. The interested reader is referred to [41] for more complete discussion of their properties.

1. Definitions of scaling or power laws

Suppose we have data with n points $\{x_1, x_2, \ldots, x_n\}$ assumed without loss of generality always to be ordered $x_1 \geqslant x_2 \geqslant \cdots \geqslant x_n$. We will say the data are *scaling* if for all $1 \leqslant k \leqslant n_s \leqslant n$, $\{x_k\}$ satisfies a power law rank vs size relationship of the form

$$kx_k^{\alpha} \approx c, \tag{A1}$$

where $c>0$ and $\alpha>0$ are constants, and where n_s determines the range of scaling [33].

Note that this description of scaling is general, in the sense that it applies to any data without regard to how it is generated and without reference to any underlying probability distributions or ensembles. A random variable X or its corresponding distribution function $F(x)$ is said to follow a *power law* (or is *scaling*) with index $\alpha>0$ if, as $x \to \infty$,

$$P[X \geqslant x] = 1 - F(x) \approx cx^{-\alpha}, \tag{A2}$$

for some constant $0<c<\infty$ and a *tail index* $\alpha>0$. That is, the cumulative probability $P[X \geqslant x]$ of observing events greater than a given size x is given by $P[X \geqslant x] \approx cx^{-\alpha}$. All moments of F of order $\beta \geqslant \alpha$ are infinite. Since relationship (A2) implies $\ln(P[X \geqslant x]) \approx \ln(c) - \alpha \ln(x)$, log-log plots of x versus $1-F(x)$ yield approximately straight lines of slope $-\alpha$, at least for large x. Well known examples of power law distributions include the Pareto distributions of the first and second kind [42]. In contrast, *exponential distributions* (i.e., $P[X \geqslant x] = e^{-\lambda x}$ result in approximately straight lines on semilogarithmic plots.

Power law distributions are called scaling distributions because the sole response to conditioning is a change in scale; that is, if X follows a power law with index α and $x \geqslant w$, the conditional distribution of X given that $X \geqslant w$ satisfies

$$P[X \geqslant x | X > w] = \frac{P[X \geqslant x]}{P[X \geqslant w]} \approx c_1 x^{-\alpha}, \tag{A3}$$

which—at least for large values of x—is identical to the (unconditional) distribution $P[X \geqslant x]$, except for a change in scale. In contrast, the exponential distribution gives

$$P(X \geqslant x | X \geqslant w) = e^{-\lambda(x-w)}. \tag{A4}$$

That is, the conditional distribution is also identical to the (unconditional) distribution, except for a change of location rather than scale.

2. More normal than "normal"

Gaussian distributions are universally viewed as "normal," mainly due to the well known central limit theorem (CLT). In particular, the ubiquity of Gaussians is largely attributed to the fact that they are invariant and attractors under aggregation of summands, required only to be independent and identically distributed and have finite variance [43]. Another convenient aspect of Gaussians is that they are completely specified by mean and variance, and the CLT justifies using these statistics whenever their estimates robustly converge, even when the data could not possibly be Gaussian. For example, much data can only take positive values (e.g., node degrees) or have hard upper bounds but can still be treated as Gaussian. It is understood that this approximation would need refinement if additional statistics or tail behaviors are of interest. Exponential distributions have their own set of invariance properties (e.g., conditional expectation) that make them attractive models in some cases.

Mathematically, perhaps the key result is that the CLT has a generalization that relaxes the finite variance (e.g., finite *CV*) assumption, allows for high variability data, and yields *stable laws* in the limit [41]. Stable laws are always scaling, with $0<\alpha<2$, except for the special case of $\alpha=2$ which gives Gaussians and corresponds to the finite variance or low variability case. From an unbiased mathematical view, the most salient features of scaling are this and additional strong invariance properties (e.g., to marginalization, mixtures, maximization), and the ease with which scaling is generated by a variety of mechanisms [33,44]. Combined with the abundant high variability in real world data, these features suggest that scaling distributions are in a sense more "normal" than Gaussians and that they are convenient and parsimonious models for high variability data in as strong a sense as Gaussians or exponentials are for low variability data.

APPENDIX B: A DERIVATION OF LATTICE ANIMAL NUMBERS g_s FOR PERCOLATION

For notational purposes define Δs_j as follows:

$$\Delta s_1 = s - s_1,$$

$$\Delta s_2 = \Delta s_1 - s_2 = s - s_1 - s_2,$$

$$\Delta s_j = \Delta s_{j-1} - s_j = s - s_1 - s_2 - \cdots - s_j.$$

Now define h_s as the number of ways to make a cluster *sub-branch* of size s on a Bethe lattice of coordination number Z. Then h_s satisfies the recursion relation

$$h_{s+1} = \sum_{s_1=0}^{s} \sum_{s_2=0}^{\Delta s_1} \cdots \sum_{s_{Z-2}=0}^{\Delta s_{Z-3}} h_{s_1} h_{s_2} \cdots h_{s_{Z-2}} h_{\Delta s_{Z-2}} \tag{B1}$$

with $h_0 \equiv 1$. Define $\tilde{g}_s$ as the number of ways to make a *whole* cluster of size s. $\tilde{g}_s$ is related to h_s by

$$\tilde{g}_{s+1} = \sum_{s_1=0}^{s} \sum_{s_2=0}^{\Delta s_1} \cdots \sum_{s_{Z-1}=0}^{\Delta s_{Z-2}} h_{s_1} h_{s_2} \cdots h_{s_{Z-1}} h_{\Delta s_{Z-1}}. \tag{B2}$$

Note that $\tilde{g}_s$ is not exactly what we want because it is not translation invariant. For example, in calculating $\tilde{g}_s$, we counted a cluster with one sub-branch of size 2 as being distinct from a cluster with two sub-branches of size 1. However, the only difference is that in the first case the "origin" is on one side of the cluster, and in the second case it is in the middle. This is easily resolved by dividing by the number of possible origins:

$$g_s = \tilde{g}_s/s. \tag{B3}$$

To solve these equations, define the generating functions

$$H(x) \equiv \sum_{s=0}^{\infty} h_s x^s \tag{B4}$$

and

$$\tilde{G}(x) \equiv \sum_{s=0}^{\infty} \tilde{g}_s x^s. \tag{B5}$$

The idea is to use the generating functions to get rid of the sums. The resulting equations will be algebraic, and much easier to analyze and solve. To do this, recursively make use of the following convolution:

$$\left(\sum_{s=0}^{\infty} a_s x^s\right)\left(\sum_{s=0}^{\infty} b_s x^s\right) = \sum_{s=0}^{\infty} c_s x^s \tag{B6}$$

where

$$c_s \equiv \sum_{s1=0}^{s} a_{s_1} b_{s-s_1}. \tag{B7}$$

In particular, note that

$$\left(\sum_{s=0}^{\infty} a_s x^s\right)^2 \tag{B8}$$

has a convolved series of

$$c_{2,s} = \sum_{s_1=0}^{s} a_{s_1} a_{s-s_1} = \sum_{s_1=0}^{s} a_{s_1} a_{\Delta s_1}, \tag{B9}$$

and

$$\left(\sum_{s=0}^{\infty} a_s x^s\right)^3 = \left(\sum_{s=0}^{\infty} a_s x^s\right)\left(\sum_{s=0}^{\infty} c_{2,s} x^s\right) \tag{B10}$$

has a convolved series of

$$c_{3,s} = \sum_{s_1=0}^{s} a_{s_1} c_{2,s-s_1} = \sum_{s_1=0}^{s} \sum_{s_2=0}^{s-s_1} a_{s_1} a_{s_2} a_{s-s1-s2} = \sum_{s_1=0}^{s} \sum_{s_2=0}^{\Delta s_1} a_{s_1} a_{s_2} a_{\Delta s_2} \tag{B11}$$

etc. Using this, we can simplify Eq. (B1). First multiply Eq. (B1) by x^s and sum over s:

$$\sum_{s=0}^{\infty} h_{s+1} x^s = \sum_{s=0}^{\infty} \sum_{s_1=0}^{s} \sum_{s_2=0}^{\Delta s_1} \cdots \sum_{s_{Z-2}=0}^{\Delta s_{Z-3}} h_{s_1} h_{s_2} \cdots h_{s_{Z-2}} h_{\Delta s_{Z-2}} x^s. \tag{B12}$$

Now notice that the right hand side is exactly the same form as the previous convolved series, i.e.,

$$\sum_{s=0}^{\infty} \sum_{s_1=0}^{s} \sum_{s_2=0}^{\Delta s_1} \cdots \sum_{s_{Z-2}=0}^{\Delta s_{Z-3}} h_{s_1} h_{s_2} \cdots h_{s_{Z-2}} h_{\Delta s_{Z-2}} x^s = \left(\sum_{s=0}^{\infty} h_s x^s\right)^{Z-1}. \tag{B13}$$

Remembering that $h_0 = 1$, the left side is a simple scaling:

$$\sum_{s=0}^{\infty} h_{s+1} x^s = \frac{1}{x}\left(-1 + \sum_{s=0}^{\infty} h_s x^s\right). \tag{B14}$$

So substituting in the definition of $H(x)$,

$$\frac{1}{x}[H(x) - 1] = H(x)^{Z-1} \tag{B15}$$

and similarly, from Eq. (B2),

$$\frac{1}{x}[\tilde{G}(x) - 1] = H(x)^Z. \tag{B16}$$

This can be arranged into the following convenient form:

$$H(x) = w + x\Phi(H(x)), \tag{B17}$$

where $\Phi(H) \equiv H^{Z-1}$ and $w = 1$. Then

$$\tilde{G}(x) = 1 + xH(x)^Z = 1 + H(x)xH(x)^{Z-1}. \tag{B18}$$

Substituting Eq. (B15),

$$\tilde{G}(x) = 1 + H(x)x\frac{1}{x}[H(x) - 1]. \tag{B19}$$

Thus we can write

$$\tilde{G}(x) = F(H(x)) \quad \text{where } F(H) \equiv 1 + H(H-1). \tag{B20}$$

From this, we could easily find a Taylor series for x in powers of $H(x)$, but we want to go the other way—we want a Taylor series for $H(x)$ in powers of x. Fortunately, the Lagrange inversion theorem [45] can do even better: it provides the following Taylor series for $\tilde{G}(x)$:

$$\tilde{G}(x) = F(w) + \sum_{s=1}^{\infty} x^s \frac{1}{s!}\frac{d^{s-1}}{dH^{s-1}}[\Phi(H)^s F'(H)]|_{H=w}. \tag{B21}$$

Since $\tilde{g}_s$ are the coefficients of x^s in $\tilde{G}(x)$,

$$\tilde{g}_s = \frac{1}{s!}\frac{d^{s-1}}{dH^{s-1}}[\Phi(H)F'(H)]|_{H=w}. \tag{B22}$$

Substituting the values of Φ and F,

$$\tilde{g}_s = \frac{1}{s!}\frac{d^{s-1}}{dH^{s-1}}[H^{(Z-1)s} \times (2H-1)]|_{H=w=1}. \tag{B23}$$

Using the product rule repeatedly,

$$\tilde{g}_s = \frac{1}{s!}(2H-1)\frac{d^{s-1}}{dH^{s-1}}H^{(Z-1)s} + \frac{2(s-1)}{s!}\frac{d^{s-2}}{dH^{s-2}}H^{(Z-1)s}|_{H=1}. \tag{B24}$$

Taking the derivatives and substituting $H=1$,

$$= \frac{1}{s!}\frac{[(Z-1)s]!}{[(Z-1)s+1]!} + \frac{2(s-1)}{s!}\frac{[(Z-1)s]!}{[(Z-1)s+2]!}, \quad \text{(B25)}$$

and a bit of algebra yields

$$\tilde{g}_s = Z\frac{[(Z-1)s]!}{[(Z-2)s+2]!(s-1)!}, \quad \text{(B26)}$$

and hence

$$g_s = Z\frac{[(Z-1)s]!}{[(Z-2)s+2]!s!}. \quad \text{(B27)}$$

[1] Mark Newman, Contemp. Phys. **46**, 323 (2005).
[2] George Kingsley Zipf, Harvard Studies Classical Philol. **15**, 1 (1929).
[3] Nigel Goldenfeld, *Lectures on Phase Transitions and the Renormalization Group* (Perseus Publishing, Reading, MA, 1992).
[4] Marc Mézard and Giorgio Parisi, J. Stat. Phys. **111**, 1 (2003).
[5] Per Bak, *How Nature Works* (Springer-Verlag, New York, 1996).
[6] Deepak Dhar and S. N. Majumdar, J. Phys. A **23**, 4333 (1990).
[7] S. E. Luria and M. Delbrück, Genetics **28**, 491 (1943).
[8] J. Vannimenus, B. Nickel, and V. Hakim, Phys. Rev. B **30**, 391 (1984).
[9] M. B. Hastings and Thomas C. Halsey, Europhys. Lett. **55**, 679 (2001).
[10] Albert-László Barabási, in *Handbook of Graphs and Networks*, edited by Stefan Bornholdt and Heinz Georg Schuster (Wiley-VCH, Weinheim, Germany, 2003).
[11] A.-L. Barabási, *Linked: The New Science of Networks* (Perseus Publishing, Cambridge, MA, 2002).
[12] A.-L. Barabási and R. Albert, Science **286**, 509 (1999).
[13] J. M. Carlson and J. Doyle, Phys. Rev. E **60**, 1412 (1999).
[14] J. M. Carlson and J. Doyle, Phys. Rev. Lett. **84**, 2529 (2000).
[15] J. Doyle and J. M. Carlson, Phys. Rev. Lett. **84**, 5656 (2000).
[16] T. Zhou and J. M. Carlson, Phys. Rev. E **62**, 3197 (2001).
[17] C. Robert, J. M. Carlson, and J. Doyle, Phys. Rev. E **63**, 056122 (2001).
[18] J. M. Carlson and J. Doyle, Proc. Natl. Acad. Sci. U.S.A. **99**, 2538 (2002).
[19] T. Zhou, J. M. Carlson, and J. Doyle, Proc. Natl. Acad. Sci. U.S.A. **99**, 2049 (2002).
[20] D. Reynolds, J. M. Carlson, and J. Doyle, Phys. Rev. E **66**, 016108 (2002).
[21] T. Zhou, J. M. Carlson, and J. Doyle, J. Theor. Biol. **236**, 438 (2005).
[22] G. Yule, Philos. Trans. R. Soc. London, Ser. B **213**, 21 (1925).
[23] M. Kleiber, Hilgardia **6**, 315 (1932).
[24] M. Kleiber, *The Fire of Life* (John Wiley and Sons, New York, 1961).
[25] M. E. Crovella and A. Bestaveros, IEEE/ACM Trans. Netw. **5**, 835 (1997).
[26] Dietrich Stauffer and Amnon Aharony, *Introduction to Percolation Theory*, 2nd ed. (Taylor & Francis, Philadelphia, 1994).
[27] R. Albert and A.-L. Barabási, Rev. Mod. Phys. **74**, 47 (2002).
[28] R. Albert, H. Jeong, and A.-L. Barabási, Nature (London) **401**, 130 (1999).
[29] R. Albert, H. Jeong, and A.-L. Barabási, Nature (London) **406**, 378 (2000).
[30] A.-L. Barabási, R. Albert, and H. Jeong, Physica A **272**, 173 (1999).
[31] A.-L. Barabási and E. Bonabeau, Sci. Am. **288**, 60 (2003).
[32] A.-L. Barabási and Z. N. Oltvai, Nat. Rev. Genet. **5**, 101 (2004).
[33] Benoit B. Mandelbrot, *Fractals and Scaling in Finance* (Springer-Verlag, New York, 1997).
[34] K. Chen, P. Bak, and M. H. Jensen, Phys. Lett. A **149**, 207 (1990).
[35] B. Drossel and F. Schwabl, Phys. Rev. Lett. **69**, 1629 (1992).
[36] B. D. Malamud, G. Morein, and D. L. Turcotte, Science **281**, 1840 (1998).
[37] M. Manning, J. M. Carlson, and Phys. Rev. E **72**, 016105 (2005).
[38] H. A. Simon, Biometrica **42**, 425 (1955).
[39] M. Mitzenmacher, Internet Math. **1**, 226 (2003).
[40] P. Ball, *Critical Mass: How One Thing Leads to Another* (Farrar, Straus and Giroux, New York, 2004).
[41] G. Samorodnitsky and M. S. Taqqu, *Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance* (Chapman and Hall, New York, 1994).
[42] N. L. Johnson, S. Kotz, and N. Balakrishnan, *Continuous Univariate Distributions*, 2nd ed. (John Wiley & Sons, New York, 1994), Vol. 1.
[43] W. Feller, *An Introduction to Probability Theory and Its Applications* (John Wiley & Sons, New York, 1971), Vol. 2.
[44] W. Willinger, D. Alderson, J. C. Doyle, and L. Li, in Proceedings of the 2004 Winter Simulation Conference, edited by R. G. Ingalls, M. D. Rossetti, J. S. Smith, and B. A. Peters (IEEE, Piscataway, NJ, 2004).
[45] Eric W. Weisstein, http://mathworld.wolfram.com/LagrangeInversion Theorem.html