

Attractor and basin entropies of random Boolean networks under asynchronous stochastic update

Amer Shreim,¹ Andrew Berdahl,^{1,2} Florian Greil,^{3,*} Jörn Davidsen,¹ and Maya Paczuski¹

¹*Complexity Science Group, Department of Physics and Astronomy, University of Calgary, Calgary, Alberta, Canada T2N 1N4*

²*Department of Ecology & Evolutionary Biology, Princeton University, Princeton, New Jersey 08544, USA*

³*Arbeitsgruppe Komplexe Systeme, Institut für Festkörperphysik, Technische Universität Darmstadt, D-64289 Darmstadt, Germany*

(Received 8 March 2010; revised manuscript received 19 July 2010; published 16 September 2010)

We introduce a method to study random Boolean networks with asynchronous stochastic update. Each node in the state space network starts with equal occupation probability, which then evolves to a steady state. Attractors and the sizes of their basins are determined by the nodes left occupied at late times. As for synchronous update, the basin entropy grows with system size only for critical networks. We determine analytically the distribution for the number of attractors and basin sizes for networks with connectivity $K=1$. These differ from the case of synchronous update for all K .

DOI: [10.1103/PhysRevE.82.035102](https://doi.org/10.1103/PhysRevE.82.035102)

PACS number(s): 89.75.Hc, 05.45.-a, 89.75.Da, 89.75.Fb

One of the main challenges in biological sciences is to identify regulatory networks from high-throughput experiments. Relevant criteria for models include conceptual simplicity, computational tractability, and robustness to uncertainties in the data. Boolean networks are candidates for representing classes of behaviors observed in large regulatory networks (see e.g., Ref. [1]). Biochemical details like reactions rates, which are often unknown, are discarded for a simpler description in which (functional sets of) genes are “on” or “off” [2,3]. In addition to describing specific small regulatory networks [4,5], Boolean networks have been used to represent other systems such as neural [6], or evolving [7] networks.

As canonical examples of disordered systems, random Boolean networks (RBNs) [8] have attracted attention from physicists for decades [2,9]. RBNs are directed graphs consisting of N Boolean elements, where each element receives input from K distinct elements. The value of the i th element, σ_i , evolves according to a random Boolean function of its K inputs: $\sigma_i(t+1) = f_i[\sigma_{i_1}(t), \sigma_{i_2}(t), \dots, \sigma_{i_K}(t)]$, where $\sigma_{i_j}(t)$ is the value of the j th input to element i at time t . Here we choose the function f_i to be zero or one with equal probability. The functions f_i are fixed in each realization of the RBN.

In the classical RBN (CRBN), elements evolve according to a globally synchronized clock. States separate into transient states, which cannot be reached more than once under the dynamics, and attractor states, which can be reached infinitely often. Attractors form cycles in state space, which divides into nonoverlapping partitions, or basins of attraction. All states within a given basin reach the same attractor at long times. Within an annealed approximation, it was found that $K=2$ CRBNs are critical [10]: they separate ordered ($K < 2$) and chaotic $K > 2$ phases, where distances between nearby trajectories vanish or diverge, respectively. Numerous investigations have focused on the lengths and numbers of attractors and their basins [2,9,11–16]. In particular, the basin entropy, which quantifies variations in the size

of the partitions of state space, increases with N only for critical CRBNs [17].

However, many real world systems do not evolve according to a globally synchronized clock, and imposing one can lead to spurious order [18]. This leaves doubts on the applicability of CRBNs as *generic* models [19–21]. To clarify the effects of the updating scheme, we consider the opposite extreme and study asynchronous, stochastic random Boolean networks (ARBNs) where at each time step, a single randomly selected element is updated.

In fact, ARBNs and CRBNs require distinct definitions of attractors and basins. Reference [22] argued, using analytic methods, that asynchronous updating drastically reduces the average number of attractors and changes its dependence on N for $K=2$. While Boolean representations of regulatory networks with synchronous updating have reproduced some empirical results [4,23], this could be due to the fact that these particular networks have evolved robust genetic wiring that enables the organism to function in a noisy environment [5,24].

Here we show using a novel computational method that while most properties of ARBNs and CRBNs are drastically different, certain features are the same for both extremes. These are generic with respect to the dynamics. In particular the basin entropy increases with system size only for $K=2$ ARBNs. Hence it is a robust and unambiguous detector of critical behavior. Further, the distribution of attractor lengths is a power law for $K=2$ networks in both cases. We also demonstrate that all attractors in a $K=1$ ARBN have the same length, and all their basins have the same size. We use this result to derive analytically the distribution of the number of attractors and the average basin entropy for ensembles of $K=1$ ARBNs as a function of N .

An attractor in a discrete system with nondeterministic dynamics, such as an ARBN, is also a subset of all possible states that can be reached infinitely often in the long-time limit starting from a random initial condition, but attractors do not necessarily form cycles. Mathematically, an attractor is a set of states such that starting from any state within it: (a) all other states of the attractor can be reached and (b) these are the only states that can be reached. States that do not

*Present address: Climate Sciences, Alfred-Wegener-Institute, D-27570 Bremerhaven, Germany.

belong to an attractor are transient. Since transient states may form loops in the state space network (SSN), our definition is not the same as in Ref. [22]. We define the normalized size of an attractor's basin to be the chance of reaching that attractor starting from a randomly chosen initial state. Note that unlike CRBNs, an ARBN can reach different attractors starting from the same initial state, i.e., the transient states in different basins can overlap. However, different attractors are still composed of nonoverlapping subsets of states.

We find attractors and their basins using the following method: First we construct the entire SSN of $\mathcal{N}=2^N$ states for an ARBN of size N , by representing the states as nodes. A directed link, with weight $1/N$, points from a node to each of its N images reached by choosing, in turn, one element for update. An image can be the state itself or some other state with Hamming distance one away. If some among these N images are identical, we add the weights of the corresponding links. Second, our algorithm initially assigns the same occupation probability $\rho=1/\mathcal{N}$ to each node. Then the algorithm updates this probability distribution on the SSN in parallel by dividing and moving the entire occupation probability on each node along its outgoing links according to the weights. The algorithm repeats this step until the probability distribution is stationary [26]. By definition, the occupation probability ρ vanishes for transient states, but remains finite for states on attractors. We identify different attractors as subsets of these latter states that are dynamically connected [27]. The value of ρ for each state is the chance of observing that state in the long-time limit after starting the ARBN from a random initial state. The sum of ρ over all states on an attractor is its normalized basin size.

Before presenting our numerical findings, we first discuss the case $K=1$, where we derive analytical results and compare them with results from the SSN method, finding complete agreement. The structure of a $K=1$ RBN is particularly simple: Boolean elements form loops with trees rooted in them. Only four Boolean functions exist to update elements: copy, invert, force to zero and force to one. Relevant components in CRBNs determine the number and lengths of attractors as well as their basin sizes [9,25]. This is also true for ARBNs. Loops of elements with at least one forcing function become frozen eventually, as do elements on trees rooted in frozen loops. Hence, relevant components are loops that contain no forcing functions. They are even or odd. Even (odd) loops have an even (odd) number of invert functions. Their number in ARBN i is $n_{even}(i)$ [$n_{odd}(i)$].

Without loss of generality, consider that all functions on even loops are copy functions. Then the two stable states are all zeroes (00...00) and all ones (11...11). Hence a single even loop generates two point attractors. By symmetry, the two point attractors split the SSN into two equal parts. Hence, for an ARBN with n_{even} loops and no odd loops, the number of attractors is $2^{n_{even}}$. Each has a basin of size $2^{N-n_{even}}$.

Odd loops are equivalent to a loop with one invert function, the rest being copy. Assume that the loop is in state $\mathbf{S}_0=\{0\dots 0\}$ and call the element with the invert function η . \mathbf{S}_0 changes only when η is chosen for update. Then, the new state \mathbf{S}_1 does not change until the neighbor on the loop downstream from η is chosen. Following this argument, we

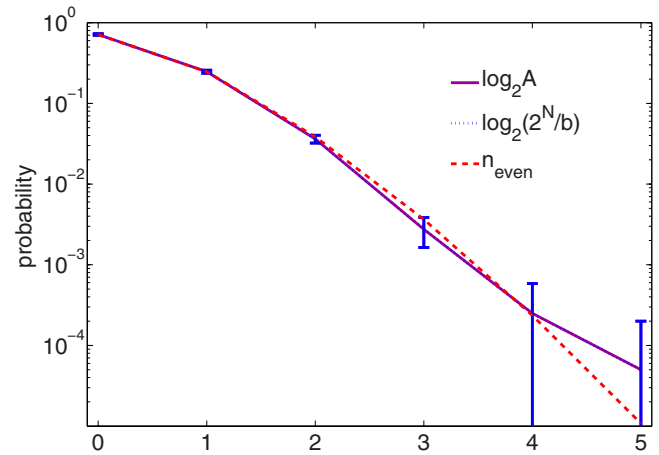


FIG. 1. (Color online) The distribution of the number of even loops $P(n_{even})$, obtained using $Q(\mathbf{m})$ —compared to the distribution of the number of attractors, $P(\log_2 A)$, and the distribution of inverse basin sizes, $P(\log_2 \frac{N}{b})$, obtained using our SSN method for $K=1$ ARBNs with $N=16$. The number of realizations is 2×10^4 and the error bars correspond to two standard deviations.

conclude that a single odd loop of length L in an ARBN has $2L$ different states it can be in. An ARBN with n_{odd} odd loops of lengths $L_1, L_2, \dots, L_{n_{odd}}$, each loop j having s_j unfrozen elements downstream of it, has one attractor with $\prod_{j=1}^{n_{odd}} 2L_j 2^{s_j}$ states. Thus for ARBN i that contains $n_{even}(i)$ and $n_{odd}(i)$ loops, the number of attractors is $A_i = 2^{n_{even}(i)}$. Since the attractor from the odd loops enters equally into all basins, each attractor α drains a basin of size $b_\alpha = 2^{N-n_{even}(i)}$. We confirmed numerically that in a given ARBN the number of states in each attractor is the same for all attractors and that all basin sizes are also equal.

We now use results of Ref. [25] to find the distribution of the number of even loops, $P(n_{even})$, over ensembles of ARBNs. These authors derive the probability, $Q(\mathbf{m})$, to observe the vector $\mathbf{m}=\{m_1, m_2, \dots\}$ relevant loops of lengths $L=\{1, 2, \dots\}$ in a $K=1$ RBN. In our case this probability reads

$$Q(\mathbf{m}) = \left(\frac{1}{2}\right)^{\hat{m}} \frac{N!}{(N-\hat{m})! N^{\hat{m}}} \frac{1}{2} (1 + \hat{m}/N) \prod_{L=1}^{\infty} \frac{L^{-m_L}}{m_L!},$$

where $\hat{m} = \sum_{L=1}^{\infty} m_L L$. The probability $P(n)$ to have n relevant loops (both even and odd) is $P(n) = Q(\mathbf{m} | \sum_L m_L = n)$. $P(n_{even})$ is obtained from $P(n)$ as

$$P(n_{even}) = \sum_{n=n_{even}}^{\infty} P(n) \binom{n}{n_{even}} \left(\frac{1}{2}\right)^n.$$

These expressions do not have closed form solutions, but can be evaluated numerically. Figure 1 shows $P(n_{even})$ evaluated from $Q(\mathbf{m})$. Noting that $P(n_{even}) = P(\log_2 A) = P(\log_2 \frac{N}{b})$, Fig. 1 also shows $P(\log_2 A)$ and $P(\log_2 \frac{N}{b})$, obtained using the numerical SSN method described previously, finding agreement.

We now consider $K \geq 1$, and distinguish two measures for the length of attractors. The first, l_α , is the actual number of states on attractor α . The second, λ_α , takes into account how

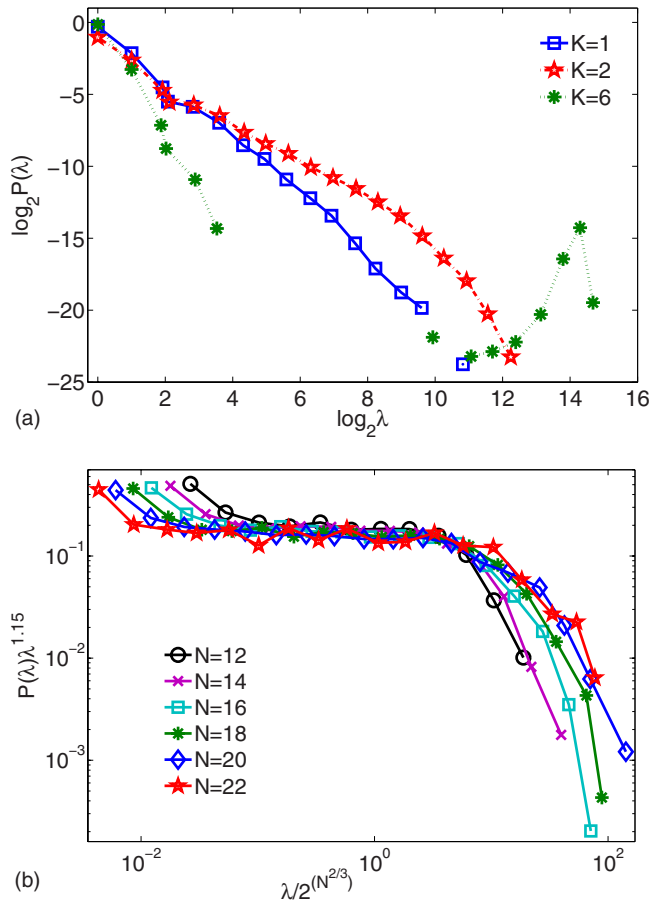


FIG. 2. (Color online) (a) $P(\lambda)$ for $K=1, 2, 6$ and $N=16$. Both $K=1$ and $K=2$ ARBNs show a broad distribution while all attractor lengths for $K=6$ are either of order unity or of order 2^N . (b) Rescaled $P(\lambda)$ vs. rescaled λ for $K=2$ and $N=12-22$. The approximate collapse is consistent with a power law decay up to a cutoff which is a stretched exponential in N .

often states are visited. Defining the stationary occupation probability of state j on attractor α as $p_{j,\alpha}$, the conditional probability of that state given that the dynamics reaches attractor α is $\varrho_{\alpha j} = \frac{p_{\alpha j}}{\sum_{j=1}^{l_\alpha} p_{\alpha j}}$. The attractor entropy, s_α , determines λ_α through $\ln \lambda_\alpha = s_\alpha = -\sum_{j=1}^{l_\alpha} \varrho_{\alpha j} \log \varrho_{\alpha j}$. If all states on attractor α are visited equally often then $\lambda_\alpha = l_\alpha$. If, on the other hand, one state dominates, $\lambda_\alpha \rightarrow 1$. We observe a strong correlation between l_α and λ_α (data not shown). Since we find clearer—although not qualitatively different—behavior using λ , we present those results below.

Figure 2(a) shows the probability density function for the length of attractors, $P(\lambda)$, for $N=16$. It is broad for $K=1$ and $K=2$. For $K=6$, λ is either of order unity or of the SSN size \mathcal{N} . This is true for all $K>2$ although subleading effects dominate for $2<K<6$, and no clear separation appears for this \mathcal{N} [28]. A rescaled $P(\lambda)$ is shown in Fig. 2(b). The cutoff in the length $\lambda_c \sim 2^{N^{2/3}}$ is suggested by the arguments in Ref. [22] and the data is consistent with a power law decay $P(\lambda) \sim \lambda^{-\tau_\lambda}$ with $\tau_\lambda = 1.15 \pm 0.05$. However, the cutoff changes shape for different N so the system is not yet in a regime where subleading corrections can be ignored.

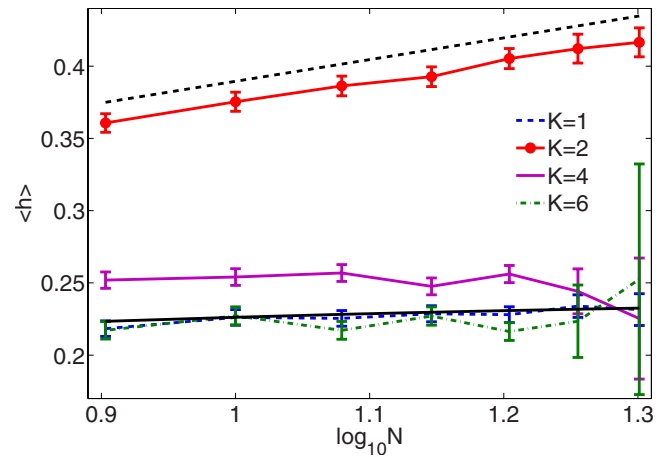


FIG. 3. (Color online) The basin entropy $\langle h \rangle$ for ARBNs with various K , and $N=8-20$ (curves with error bars). $\langle h \rangle$ grows with N only for $K=2$. The solid (black) line shows $\langle h \rangle$ for $K=1$ derived from $Q(\mathbf{m})$. The dashed (black) line has a slope of 0.15 and is a guide for the eyes.

The probability, p_α , to reach attractor α from a random initial state is the sum of all occupational probabilities, $p_\alpha = \sum_{j=1}^{l_\alpha} p_{\alpha j}$ over states on α . Its basin size, $b_\alpha = \mathcal{N} p_\alpha$. Reference [17] proposed that the average basin entropy $\langle h \rangle$, which gives a summary measure for the variation in basin sizes, is a measure of complexity and criticality in CRBNs. The basin entropy of RBN i is $h_i = -\sum_{\alpha=1}^{A_i} p_\alpha \ln p_\alpha$, and the average basin entropy over the ensemble is $\langle h \rangle = \frac{1}{R} \sum_i^R h_i$ where R is the number of realizations. Figure 3 shows the average basin entropy $\langle h \rangle$ for ARBNs with $K=1, 2, 3, 6$ and $N=8$ to 20. This figure shows that $\langle h \rangle$ grows only for $K=2$ while remaining constant for $K \neq 2$. For $K=1$, Fig. 3 also shows $\langle h \rangle$ derived from $Q(\mathbf{m})$, which agrees perfectly with the SSN method. For CRBNs [17], as well, the basin entropy was shown to grow with system size only for the critical case $K=2$ —over the same range of system sizes. This is particularly interesting since all other properties we measured, such as the number and size of attractors and their basins (data not shown), are remarkably different for synchronous and asynchronous updating.

Reference [15] showed that the distribution of attractor lengths in CRBNs depends on the sampling scheme. For CRBNs with $K>2$ the distribution obtained by counting all attractors is a power law, while that obtained by randomly sampling initial states is not. For $K=2$ both distributions are power laws. This is due to a lack of correlation between the size of an attractor and the size of its basin for $K=2$, while for $K>2$ the basin size grows linearly with attractor length. In the ARBNs studied here, the basin size varies little with attractor length for all K , so the two distributions are indistinguishable over the range of system sizes $8 \leq N \leq 20$.

As in Ref. [17], a method based on exact enumeration of state space creates a severe restriction on the system sizes that can be studied. Indeed for $K=2$ CRBNs, the mean number of attractors grows faster than any power law with system size [13], but one cannot reach this asymptotic result for CRBNs in the range of $8 \leq N \leq 20$. The growth we observed

for CRBNs is much slower (data not shown). For ARBNs, the mean number of attractors increases as a power law with system size for $K=2$ [22]. We also do not observe this asymptotic result for the range of system sizes studied (data not shown). Hence, it is not possible to extract an asymptotic growth law for the size dependence of the basin entropy for critical networks using this method in either case.

To summarize, we have developed a method to find the long-time dynamics of ARBNs. It relies on the flow of the occupation probability over the complete network of possible states until stationarity is reached. We focused on how the space of states divides into overlapping basins of attraction and measured the fluctuations of these fuzzy partitions in ensembles of ARBNs. Despite the fundamental difference between ARBNs and classical RBNs, and the fact that all other observables show major differences, we find that in

both cases the basin entropy increases with system size only for critical networks. This suggests that “superuniversal” features exist for the fluctuations in the structure of state space of RBNs—these are invariant with respect to the specific dynamics. As the dynamics for real world systems most likely lies in between these two extremes, our work suggests that the basin entropy may be a relevant and robust signifier of criticality, irrespective of the specific dynamics. Finally, we showed that all attractors in an ordered $K=1$ ARBN have the same length, and that all their basins have the same size. We used this to derive analytically the distributions of both basin entropy and the number of attractors in $K=1$ ARBNs as a function of system size.

This work was financially supported by NSERC, the Yukon Foundation, and the DAAD.

-
- [1] M. Lagomarsino, B. Bassetti, G. Castellani, and D. Remondini, *Mol. BioSys.* **5**, 335 (2009).
- [2] M. Aldana, S. Coppersmith, and L. Kadanoff, in *Perspectives and Problems in Nonlinear Science*, edited by E. Kaplan, J. Marsden, and K. Sreenivasan (Springer, New York, 2003), pp. 23–89.
- [3] S. Bornholdt, *J. R. Soc., Interface* **5**, S85 (2008).
- [4] R. Albert, *J. Cell. Sci.* **118**, 4947 (2005).
- [5] F. Li, T. Long, Y. Lu, Q. Ouyang, and C. Tang, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 4781 (2004).
- [6] M. Rosen-Zvi, A. Engel, and I. Kanter, *Phys. Rev. Lett.* **87**, 078101 (2001).
- [7] M. Paczuski, K. E. Bassler, and Á. Corral, *Phys. Rev. Lett.* **84**, 3185 (2000).
- [8] S. Kauffman, *J. Theor. Biol.* **22**, 437 (1969).
- [9] B. Drossel, in *Reviews of Nonlinear Dynamics and Complexity*, edited by H. Schuster (Wiley-VCH, Weinheim, 2008), Vol. 1.
- [10] B. Derrida and Y. Pomeau, *EPL* **1**, 45 (1986).
- [11] B. Drossel, *Phys. Rev. E* **72**, 016110 (2005).
- [12] B. Drossel, T. Mihaljev, and F. Greil, *Phys. Rev. Lett.* **94**, 088701 (2005).
- [13] B. Samuelsson and C. Troein, *Phys. Rev. Lett.* **90**, 098701 (2003).
- [14] U. Bastolla and G. Parisi, *Physica D* **98**, 1 (1996).
- [15] A. Berdahl, A. Shreim, V. Sood, M. Paczuski, and J. Davidsen, *New J. Phys.* **11**, 043024 (2009).
- [16] A. Bhattacharjya and S. Liang, *Phys. Rev. Lett.* **77**, 1644 (1996).
- [17] P. Krawitz and I. Shmulevich, *Phys. Rev. Lett.* **98**, 158701 (2007).
- [18] B. A. Huberman and N. S. Glance, *Proc. Natl. Acad. Sci. U.S.A.* **90**, 7716 (1993).
- [19] I. Harvey and T. Bossomaier, *Proceedings of the Fourth European Conference on Artificial Life* (MIT Press, Cambridge, 1997), pp. 67–75.
- [20] K. Klemm and S. Bornholdt, *Phys. Rev. E* **72**, 055101(R) (2005).
- [21] L. Glass and C. Hill, *EPL* **41**, 599 (1998).
- [22] F. Greil and B. Drossel, *Phys. Rev. Lett.* **95**, 048701 (2005).
- [23] M. I. Davidich and S. Bornholdt, *PLoS ONE* **3**, e1672 (2008).
- [24] S. Braunewell and S. Bornholdt, *J. Theor. Biol.* **245**, 638 (2007).
- [25] H. Flyvbjerg and N. Kjaer, *J. Phys. A* **21**, 1695 (1988).
- [26] In practice, our criteria for stationarity requires that for each node in the SSN, the change in its occupation probability after an update is smaller than ϵ . Here we use $\epsilon=10^{-6}/N$.
- [27] To check for consistency, we tested whether there were states identified erroneously as transient that could be reached from an attractor. This was never observed.
- [28] Note that the corresponding distribution for CRBNs, $P_a(l)$ in the notation of Ref. [15], is broad for all $K \geq 2$, exactly the opposite as found here for ARBNs.