A simple solution of the critical Kauffman model with connectivity one

T. M. A. Fink and F. C. Sheldon

London Institute for Mathematical Sciences, Royal Institution, 21 Albemarle St, London W1S 4BS, UK

(Dated: March 3, 2023)

The Kauffman model is a simple model of genetic computation which highlights the importance of critical behavior at the border of order and chaos. We present a simple proof that the number and length of attractors for the critical Kauffman model with connectivity one exhibit super-polynomial growth with the size of the network, improving on the best known bounds. Our approach is to bound the mean attractor length from above and below using elementary methods in number theory.

The Kauffman model is a simplified model of biological networks that has been widely studied [1–3]. It highlights the importance of criticality—the border between frozen and chaotic dynamics at which many biological systems appear poised [4].

The Kauffman model with connectivity one plays a special role because it is exactly solvable [1]. At the same time, new approaches to solving it reveal additional insights [2, 3], suggesting techniques for more realistic models which cannot be solved exactly. If, as is widely believed [2], all critical Kauffman models behave in a similar way, our results for the connectivity-one model may apply to other critical versions too.

In a Kauffman model with connectivity one, N nodes form a random directed network such that each node has one input, but any number of outputs. Thus the network is composed of loops and trees branching off the loops. Because the nodes in the trees are slaves to the loops, they do not contribute to the number or length of attractors, which are set solely by the m nodes in the loops. Each node is randomly assigned one of four Boolean functions: on, off, copy or invert.

In a critical Kauffman model, a perturbation to one node propagates to, on average, one other node. So in the critical model with connectivity one, all of the Boolean functions must be copy or invert [2, 3]. Because of this, all of the 2^m states of the m nodes in loops form cycles; there are no transients.

Throughout, we use the word dynamics to refer to the number and size of cycles. For a loop of size l, the dynamics depends only on the parity of the number of inverts in the loop. If the number of inverts is even, it is called an even loop, and it has cycles of length k if k divides l. If the number of inverts is odd, it is called an odd loop, and it has cycles of length k if k divides $2l$ but not l. We use the shorthand $\{l\}$ and $\{\bar{l}\}$ for even and odd loops of length l. Let $g \circ x$ denote g copies of an x-cycle. Then we can write the dynamics as $d = g_1 \circ x_1 + g_2 \circ x_2 + \ldots$, where "+" means "and". Examples are given in Table I.

Consider a collection of s loops $\{l_1, l_2, \ldots, l_s\}$, where $m = l_1 + \ldots + l_s$ is the number of nodes in loops. The dynamics of multiple loops can be deduced from the dynamics of individual loops by defining a product between them [3]:

$$(g_1 \circ x_1 + g_2 \circ x_2 + \ldots)(h_1 \circ y_1 + h_2 \circ y_2 + \ldots) = \sum_{i,j} g_i h_j \; \gcd(x_i, y_j) \circ \mathrm{lcm}(x_i, y_j). \tag{1}$$

Examples are given in the right of Table I.

We can bound the mean attractor size from above by calculating the largest attractor size. In passing we note that in general most states belong to one of the largest attractors. For example, of the $2^6 = 64$ states in $d(6)$, 54 of them belong to cycles of length 6 (see Table I). In fact the minimum fraction of states in the largest cycle is

$$\prod_{i=1}^{\infty} \left(1 - \frac{2}{2^{p_i}}\right) = 0.346.$$

Why? For even and odd loops of length odd prime p, the number of cycles of length p and $2p$ is $(2^p - 2)/p$ and $(2^p - 2)/(2p)$. When the loop length is not prime, the number of cycles is greater than $(2^l - 2)/l$ and $(2^l - 2)/(2l)$ [3]. Combining these with eq. (1), the result follows.

In the presence of odd loops, the largest cycle length is double the least common multiple of the individual loop sizes. Consider all ways of partitioning some number m into $l_1, l_2, \ldots$. For $m = 8$, for example, many of these partitions are shown in the right of Table I. What is the maximum value of the least common multiple of the partitions? This is precisely Landau's function $g(m)$ (OEIS A000793 [6]), the first few values of which are 1, 2, 3, 4, 6, 6, 12, 15, 20, 30. When m is the sum of the first primes, the values are just the product of the primes. Landau's function generalizes this notion for other values of m. In particular it is known that $g(m) \leq 2^{1.52\sqrt{m \ln m}}$ [7]. Therefore the mean attractor length $\overline{A}$ satisfies

$$\overline{A}(l_1, l_2, \ldots) < 2^{1.52\sqrt{m \ln m}+1}. \tag{2}$$

This is considerably less than previous bounds of $2^{0.53m}$ in [2] and $2^{0.5m}$ in [1].

The upper bound on the attractor length gives a lower bound on the number of attractors. Writing $d(l_1, l_2, \ldots) = \nu_1 \circ A_1 + \nu_2 \circ A_2 + \ldots$, where there are ν_i cycles of length A_i, the mean attractor length is

$$\overline{A} = \sum_i \nu_i A_i \Big/ \sum_i \nu_i. \tag{3}$$

Even loops	*Odd loops*	*Multiple positive loops*
$d(1) = 2 \circ 1$	$d(\overline{1}) = 1 \circ 2$	$d(1,7) = 4 \circ 1 + 36 \circ 7$
$d(2) = 2 \circ 1 + 1 \circ 2$	$d(\overline{2}) = 1 \circ 4$	$d(2,6) = 4 \circ 1 + 6 \circ 2 + 4 \circ 3 + 38 \circ 6$
$d(3) = 2 \circ 1 + 2 \circ 3$	$d(\overline{3}) = 1 \circ 2 + 1 \circ 6$	$d(3,5) = 4 \circ 1 + 4 \circ 3 + 12 \circ 5 + 12 \circ 15$
$d(4) = 2 \circ 1 + 1 \circ 2 + 3 \circ 4$	$d(\overline{4}) = 2 \circ 8$	$d(4,4) = 4 \circ 1 + 6 \circ 2 + 60 \circ 4$
$d(5) = 2 \circ 1 + 6 \circ 5$	$d(\overline{5}) = 1 \circ 2 + 3 \circ 10$	$d(2,3,3) = 8 \circ 1 + 4 \circ 2 + 40 \circ 3 + 20 \circ 6$
$d(6) = 2 \circ 1 + 1 \circ 2 + 2 \circ 3 + 9 \circ 6$	$d(\overline{6}) = 1 \circ 4 + 5 \circ 12$	$d(2,2,4) = 8 \circ 1 + 28 \circ 2 + 48 \circ 4$
$d(7) = 2 \circ 1 + 18 \circ 7$	$d(\overline{7}) = 1 \circ 2 + 9 \circ 14$	$d(2,2,2,2) = 16 \circ 1 + 120 \circ 2$
$d(8) = 2 \circ 1 + 1 \circ 2 + 3 \circ 4 + 30 \circ 8$	$d(\overline{8}) = 16 \circ 16$	$d(1,1,1,1,1,1,1,1) = 256 \circ 1$

TABLE I: **Number and length of cycles for single and multiple loops.** An even loop of length l has cycles of length k if k divides l. For example, $d(3)$ reads as 2 cycles of length 1 and 2 cycles of length 3. An odd loop of length l has cycles of length k if k divides $2l$ but not l. The cycle lengths of multiple loops are the least common multiples of the cycle lengths of the individual loops. For 8 nodes in loops (right), the largest cycle length is found for two prime loops: 3 and 5.

Since all 2^m states of the loop nodes belong to cycles,

$$\sum_i \nu_i A_i = 2^m.$$

Inserting this and eq. (2) into eq. (3), the total number of attractors $c(m)$ for m nodes in loops satisfies

$$c(m) = \sum_i \nu_i > 2^{m-1.52\sqrt{m \ln m}-1}.$$

Now let's re-express this in terms of the total number of nodes N, whereby m becomes a random variable. In the large N limit, the mean number of loops of length l is $\exp\left(-l^2/(2N)\right)/l$ [1]. Summing over this, the mean number of nodes in loops $\overline{m}$ is asymptotically $\sqrt{\frac{\pi}{2}N}$. Since eq. (4) is convex, by Jensen's inequality $c(N)$ grows as

$$c(N) > 2^{1.25\sqrt{N}},$$

faster than the best known bound of $2^{0.57\sqrt{N}}$ [2].

We now turn to bounding the mean attractor length from below. The attractor length is the least common multiple of the cycle lengths of the loops. Let σ be a binary vector of length L, where $\sigma_l = 1$ if at least one loop of length l occurs and $\sigma_l = 0$ otherwise. The lower bound for the mean occurs when all loops are even, so

$$\begin{aligned}\overline{A} &\geq \sum_\sigma P(\sigma)\mathrm{LCM}\left(1^{\sigma_1}, \ldots N^{\sigma_N}\right) \\ &\geq \sum_\sigma P(\sigma) \prod_{\text{prime } p \leq N} p^{\sigma_p}, \end{aligned} \tag{4}$$

where we have further bounded the least common multiple with the primorial. We introduce a maximum loop length $L \ll N$ below which the individual loops can be regarded as independent in the large N limit. From [2], for loop lengths $L < \sqrt{\epsilon N}$ with $\epsilon \ll 1$, the numbers of loops of size l are independent and Poisson distributed with mean $1/l$. Thus in the primorial we can regard each prime as independent and included in the product with probability $1 - e^{-1/p}$.

Summing over σ and expanding $e^{-1/p}$,

$$\overline{A} > \prod_{\text{prime } p \leq L} \left(1 - \tfrac{1}{p}\right) + p\left(\tfrac{1}{p} - \tfrac{1}{2p^2}\right) = 2^k \prod_{\text{prime } p \leq L} \left(1 - \tfrac{3}{4p}\right),$$

where $k = L/\ln L$ is a lower bound on the number of primes, valid for $L > 17$ [5]. A generalisation of Merten's third theorem shows that the product on the right goes to zero as $C/\ln^{3/4} L$ [5]. Using the upper bound on the loop size, we have

$$\overline{A} > C \frac{2^{\sqrt{\epsilon N}/\ln\sqrt{\epsilon N}}}{\ln^{3/4}\sqrt{\epsilon N}}. \tag{5}$$

This is a substantial improvement over $N^\alpha e^{\beta \ln^2 N}/\left(\ln(N)/\gamma\right)!$ in [2], where α, β and γ are constants, though this only becomes apparent at very large values of N.

Simplified proofs are useful because they improve our understanding of why a system behaves as it does. Our proofs that the number and length of attractors exhibit superpolynomial growth are much simpler than previous efforts [1, 2]. They also give tighter bounds. The scaling of these quantities is important in understanding how the Kauffman model might represent physical systems, such as metabolic networks or genetic computation. The techniques used in simplified proofs are more likely to be transferable, and we anticipate that our number-theoretic approach may be applicable to other critical Kauffman models.

[1] H. Flyvbjerg, N. J. Kjaer, J Phys A **21**, 1695 (1988).
[2] B. Drossel, T. Mihaljev, F. Greil, Phys Rev Lett **94**, 088701 (2005).
[3] T. Fink, arXiv:2302.05314.
[4] M. A. Muñoz, Rev Mod Phys **90**, 031001 (2018).
[5] G. Tenenbaum, *Introduction to Analytic and Probabilistic Number Theory* (Cambridge University Press, Cambridge, 1995).
[6] N. J. A. Sloane, editor, On-Line Encyclopedia of Integer Sequences, published electronically at https://oeis.org.
[7] J.-P. Massias, Ann Fac Sci Toulouse Math **6**, 269 (1984).