Magic Number 7 ± 2 in Networks of Threshold Dynamics

Shuji Ishihara[1,*] and Kunihiko Kaneko[1,2]

[1]*Department of Pure and Applied Sciences, University of Tokyo, Komaba, Meguro-ku, Tokyo 153-8902, Japan*
[2]*ERATO Complex Systems Biology Project, Japan Science and Technology Corporation, Shibuya, Tokyo 150-0001, Japan*
(Received 20 September 2004; published 9 February 2005)

Information processing by random feed-forward networks consisting of units with a sigmoidal input-output response is studied by focusing on the dependence of its outputs on the number of parallel paths M. It is found that the system leads to a combination of on-off outputs when $M \lesssim 7$, while for $M \gtrsim 7$, chaotic dynamics arises, resulting in a continuous distribution of outputs. This universality of the critical number $M \sim 7$ is explained by a combinatorial explosion, i.e., the dominance of factorial over exponential increase. The relevance of the result to biological problems is briefly discussed.

DOI: 10.1103/PhysRevLett.94.058102 PACS numbers: 87.19.Bb, 05.40.–a, 05.45.Gg, 87.10.+e

Information processing (IP) in biological systems is often carried out by elements that show threshold-type (sigmoid) input-output (IO) behaviors. For example, the expression of a gene is determined by an on-off switch for its transcription factors. Another example is the signal transduction system in a cell where the enzymatic reaction often displays a sigmoid output when responding to an external stimulus like the abundance of an input chemical [1]. One of the most well-studied examples of sigmoid IO relationships can be found in neural networks, where the output of each neuron depends upon the input from other neurons or a receptive field [2].

In these biological networks, the connections among elements are often entangled. Cross talk in signal transduction has recently been observed for several systems [1], while enzymatic reactions are generally entangled as well. The connections in neural networks are known to be complex. Besides complexity, biological networks can also display cascade type structures leading from the external input to the final output. In signal transduction such cascades are encountered, while layered networks have been discussed as an idealization of biological neural networks as layered neural network (LNN). Hence, the study of layered network is generally important. The information processing in such systems is discussed by judging whether distinct attracting points (or sets) are reached through the dynamics in successive layers, depending on the input. In a layered network system, e.g., the attracting set is the state of the output layer.

In a LNN, the more the number of degrees of freedom to be processed increases, the more mutual interference can occur thus increasing the complexity. Consequently, the IP ability of the network depends on the number of processed degrees of freedoms. In this Letter, we discuss this number dependence and explore some universal properties of entangled networks with sigmoid units.

In order to investigate the question raised above, we adopt a cascade perceptron as an abstract model of a random sigmoid response[3–5]. Here we consider feed-forward network dynamics without feedback loop for simplicity, and adopt LNN with random couplings as the model [4,5]. In the model each layer l is composed of M elements and all elements are regulated by the elements in the preceding layer:

$$x_i^l = \tanh\left(\frac{\beta}{\sqrt{M}}\sum_k^M \epsilon_{ik}^l x_k^{l-1} - \theta_i^l\right), \tag{1}$$

where x_i^l represents the state of the ith element of the lth layer. θ_i^l is the threshold value for x_i^l to be "excitatory." Unless otherwise stated, we set $\theta_i^l = 0$ as this specific choice is not important for the later discussion. The l-dependent coupling terms ϵ_{ik}^l are chosen randomly from a Gaussian distribution with standard deviation 1.0 . The parameter β normalized by $\sqrt{M}$ determines the steepness of the sigmoid function. As β approaches 0, Eq. (1) approaches a constant function with $x_i^l = 0$ (or $x_i^l = -\theta_i^l$). On the other hand, as β increases, Eq. (1) approaches a step function such that in almost all cases x_i^l becomes either -1 or 1, and each element effectively has just two states. For the medium range of β, the IO relationship is smooth, which, as will be shown later, may lead to complex dynamics. Note that if all the thresholds $\theta_i^l = 0$, the change of sign $x \to -x$ preserves the equations of the system so that the solutions for Eq. (1) are symmetric.

For the information processing stages carried out in each layer, $\vec{x}^l = F_l(\vec{x}^{l-1})$ holds where $\vec{x}^l \equiv (x_1^l, x_2^l, \cdots, x_M^l)$ and F_l is the processing dynamics carried out in the lth layer. We set the values of the 0th layer as the inputs to be processed. If a succeeding layer is regarded as a next time step, the present system can be interpreted as a random dynamical system [6] where the processing corresponds to temporal evolution. We take various inputs (corresponding to a *set of inputs*) randomly chosen such that $x_i^0 \in [-1, 1]$ for each i, and compute numerically the evolution of Eq. (1).

Let us first discuss the qualitative behavior of Eq. (1). For $\theta_i^l = 0$, x_i^l the outputs converge to 0, irrespectively of the inputs if $\beta < 1$, while they approach either 1 or -1 when β is large. For middle range values of β, outputs may take values between -1 and 1, depending on the input. In this regime, outputs are often sensitive to changes of the inputs, and indeed orbital instability exists in the evolution through the layers. The degree of this instability depends on the number of parallel paths and on interference. For sufficiently small $M(\lesssim 7)$, the IP is stable, in the sense that the x_i^l in the output layer only assume a few distinct values, depending on the input values. On the other hand, if M is large ($\gtrsim 7$), such convergence is not common. We have computed a histogram of the output values $P(x_1^l)$ sampled over 5.0×10^5 randomly chosen inputs. As shown in Fig. 1, there are clearly two peaks at $x = \pm 1$ for $M = 6$ and $P(x) \approx 0$ for $x \neq 1$, while for $M = 8$, the distribution is broad.

The scattering in the values x_i^l of the attractor for the latter case is due to chaotic dynamics in Eq. (1), where stretching and folding in phase space appear [7]. Figures 2(a) and 2(b) show values of x^l projected onto the (x_1^l, x_2^l) plane. In the plot we take $l = 30$ and 10^5 inputs given at $l = 0$. For $M = 6$ [Fig. 2(a)] each x_i^l at $l = 30$ is localized within a small volume of the total phase space. On the other hand, for $M = 8$ [Fig. 2(b)], one can see folding and stretching, and the scattering of points throughout phase space. With these chaotic dynamics, tiny differences in the input values are amplified making clear separation of inputs impossible.

The above simulations are carried out with $\beta = 3.0$, but this stretching and folding process is observed as long as $\beta(>1)$. To obtain insight into the dependence on β, values of $x_1^{l=30}$ for 800 inputs are plotted as a function of β in Fig. 3. For $M = 6$, they converge to a few points for a large portion of β, while for [Fig. 3(b)] $M = 8$ they do not.

These numerical results suggest that a critical number of parallel paths M_c exist, beyond which chaotic dynamics is inevitable, and that M_c is around 7 for a wide range of β [8,11]. We confirm this critical number by computing several characteristic quantities for the model Eq. (1).

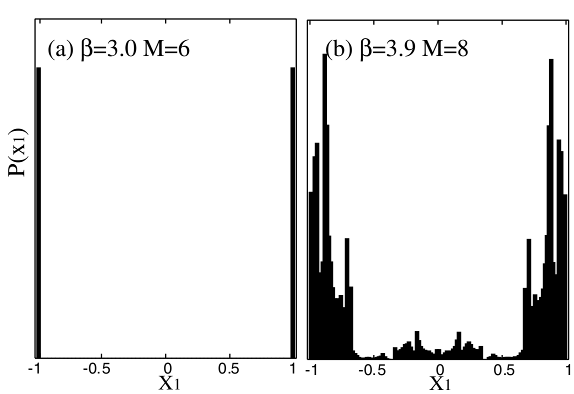

FIG. 1. Distribution of $P(x_1^l)$ at layer $l = 30$, obtained over 5.0×10^5 inputs with a homogeneous distribution over $[-1, 1]$. The parameter $\beta = 3.0$ and the threshold θ is set to 0. The histogram is plotted by using a bin size $2.0/128$. (a) for $M = 6$; (b) for $M = 8$.

First, we plot the fraction of bins for which $P(x_1^l)$ is not zero in each layer in Fig. 4(a). Here $P(x_1^l)$ is computed over 5.0×10^5 inputs by taking a bin size of $2.0/128$. For $M \leq 7$ the fraction becomes smaller from layer to layer, while for $M \geq 8$ the fraction is almost one and does not decrease much for successive layers. For $M \leq 7$, the output points are well separated by the sigmoid function, while they are scattered over the whole range of values $[-1, 1]$ for $M \geq 8$. The data for Fig. 4(a) are obtained for a fixed threshold $\theta_i = 0$, but the conclusion does not change even when the thresholds are distributed, as shown in Fig. 4(b) where $\theta_i \in [0, 0.5]$. These behaviors are also invariant against changes in β, as long as it is sufficiently larger than 1 but not that large for the tanh function to effectively become a step function. Hence the critical number $M_c \approx 7$ is rather general, without dependencies on the details of the model.

Second, we have computed the degree of orbital instability in the chaotic dynamics, i.e., the sensitivity on input values. By regarding a layer as a time step in a dynamical system, the sensitivity is computed by the Lyapunov exponent of the random map Eq. (1) as follows [9]:

$$\lambda_{\max} = \max_{\vec{x}, |\vec{x}|=1} \frac{1}{l} \ln|J^l J^{l-1} \cdots J^1 \vec{x}|, \tag{2}$$

where J^l is the Jacobian matrix of Eq. (1), $J_{ik}^l \equiv \frac{\partial x_i^l}{\partial x_k^{l-1}}$, so that $\delta x_j^l = \sum_k J_{ik}^l (x^{l-1}) \delta x_k^{l-1}$. The fraction of the network having positive exponents $\lambda_{\max}$ is plotted in Fig. 5 for the following three cases: $\theta_i^l = 0$ with a Gaussian distribution for ϵ_{ik}^l, $\theta_i^l = 0$ with a uniform distribution for $\epsilon_{ik}^l \in [-1, 1]$, and distributed thresholds $\theta_i^l \in [0, 0.5]$ with a Gaussian distribution for ϵ_{ik}^l. For all of the three cases, the fraction of networks with chaotic behavior drastically increases around $M = 7$.

Loss of separability of inputs around the number 7 due to chaotic dynamics is not limited to the model investigated above. We have also investigated some other models con-

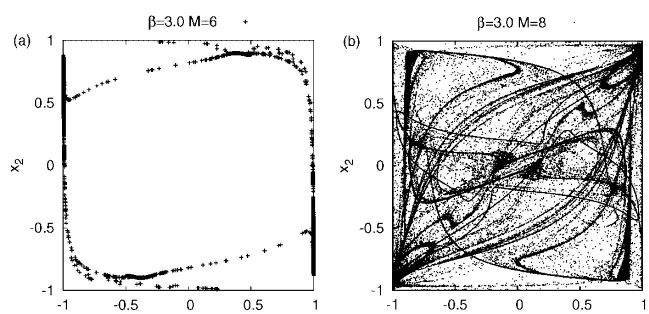

FIG. 2. Snapshots of (x_1^l, x_2^l) at layer $l = 30$, plotted as a projection of $\vec{x}^l$ onto the two-dimensional plane (x_1, x_2). The data obtained from 10^5 randomly chosen inputs are overlaid with $\beta = 3.0$. (a) $M = 6$. The density of points is quite high around $(\pm 1, \pm 1)$. (b) $M = 8$. Multiple folding and stretching processes are detected, while the points are scattered over the plane.

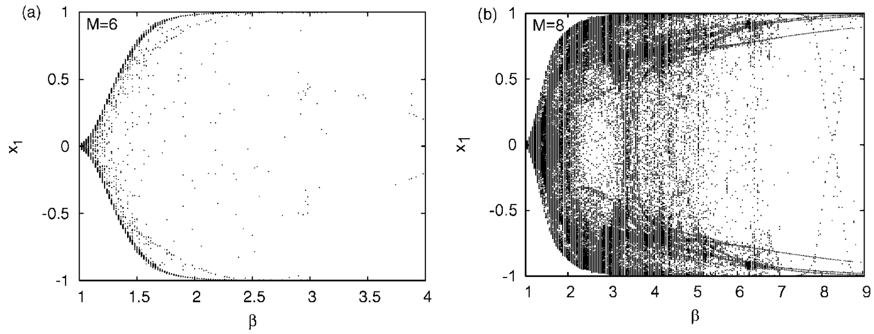

FIG. 3. Bifurcation diagram of $x_1^{l=30}$ against β. x_1^l at layer $l = 30$ is plotted over 800 different inputs for each β. For $\beta(\lesssim 1.0)$, x^l converge to $x_1 = 0$. (a) $M = 6$. For most inputs, $x_1^{l=30}$ converges to ± 1 if β is sufficiently larger than 1. (b) For $M = 8$, $x_1^{l=30}$ is scattered over $[-1, 1]$.

sisting of units with threshold dynamics (of Michaelis-Menten's form for enzymatic reactions) that are randomly connected in a cascade [12]. The same behavior with the same critical number 7 is obtained. On the other hand, it is also interesting to note that Milnor attractors that collide with their basin boundary are dominant for globally coupled dynamical systems with more than 7 ± 2 degrees of freedom [13–15].

Why, then, is the critical number 7 (or 7 ± 2) so universal? In [13], one of the authors (K. K.) discussed the possibility that the combinatorial explosion of the basin boundaries due to chaotic dynamics is relevant to this critical number (i.e., the faster increase of $N!$ over 2^N). This combinatorial argument can be extended to the present problem.

We do so by considering the origin of the folding process. In order to see the effect of entanglement, we study the input-output relationship of $x_1^0 \to x_1^2$ of a two-layer system [16] by fixing the inputs of $x_2^0, \cdots, x_M^0$. Then output x_1^2 is given as a function of x_1^0; $x_1^2 = \tanh[\sum_j \sigma_j(x_1^0 - \nu_j)]$ where $\sigma_j(u) = \tilde{\beta}\epsilon_{1j}^2 \tanh(\tilde{\beta}\epsilon_{j1}^1 u)$ and $\nu_j = (\epsilon_{j1}^l)^{-1} \times \sum_{k=2}^M \epsilon_{jk}^1 x_k^0$. Here it is assumed that $\tilde{\beta} \equiv \beta/\sqrt{M}$ is large, and that $\tanh(x)$ is close to a step function. Note that there are N paths via the middle layer elements where x_j^1 switches between the values -1 and 1 as x_1^0 crosses the "threshold" ν_j. One can then renumber the index $j = 1, \cdots, M$ such that $-1 < \nu_1 < \nu_2 < \cdots < \nu_M < 1$. With this ordering, if σ_j is positive and σ_{j+1} is negative, the one-dimensional mapping $x_1^0 \to x_1^2$ has a single hump at $\nu_j < x < \nu_{j+1}$ implying a folding process as in the logistic map. Then, if the sign of σ_j alternates for successive j, the above function switches between -1 and 1 at ν_j M times as x_1^0 is increased. The one-dimensional mapping from the input x_1^0 to the output x_1^2 is thus subject to this folding process everywhere in $-1 < x < 1$. Since ϵ_{1k}^2 can take positive or negative values with equal probability, the probability to have full folding decreases proportionally to 2^{-M}.

The estimate given so far is for fixed inputs of $x_2, x_3, \cdots, x_M$. By changing these input values, the ordering of ν_j changes accordingly (for the original index without reordering) and hence there are in total $(M - 1)!$ possible orderings. Therefore, roughly speaking, the input-output relationship has full up-down switches for some input values $x_2, x_3, \cdots, x_M$ when $(M - 1)!2^{-M}$ exceeds 1. In this case, at every layer, for any element, the folding occurs fully for some inputs, and the folding process covers most of phase space. Even though this argument is quite rough, it is still possible to presume that when $(M - 1)!$ exceeds the order of 2^M, the chaotic dynamics replaces the separation by the threshold function. Note that this facto-

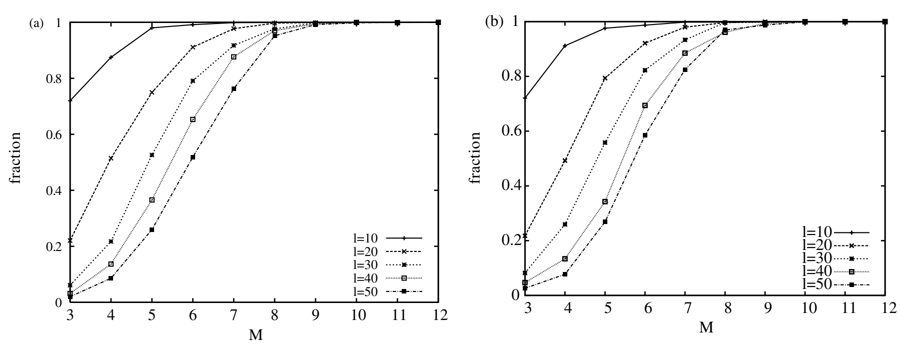

FIG. 4. The average fraction of the value x such that $P(x_1^l = x) > 0$, plotted as a function of M. The histogram is computed for 10^5 inputs over 200 networks (i.e., with different choices of ϵ_{ik}^l), with a bin size 2.0/128. The fraction at layer $l = 10, 20, 30, 40, 50$ is plotted. (a) With threshold $\theta_i^l = 0$ and (b) with distributed threshold $\theta_i \in [0, 0.5]$. The parameter β is fixed at 3.0.

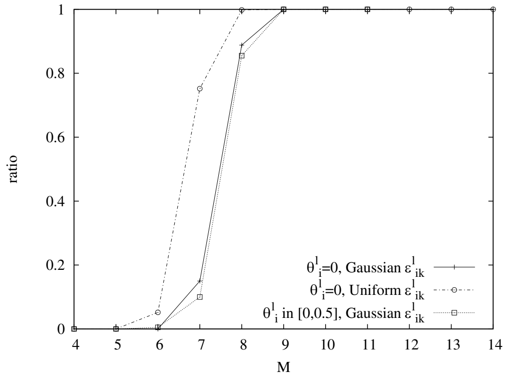

FIG. 5. The ratio of networks that shows "chaotic" behavior for some values of β, plotted as a function of M. 500 networks are chosen for ϵ^l_{ik}. The Lyapunov exponent is computed by using J^l over 500 steps (layers). The behavior is judged numerically as chaotic if the exponent is larger than 0.0. The ratio is computed for the following three cases: the threshold $\theta^l_i = 0$ with a Gaussian distribution of ϵ^l_{ik} (cross), distributed thresholds $\theta_{ij} \in [0, 0.5]$ with a Gaussian distribution of ϵ^l_{ik} (box), and $\theta^l_i = 0$ with a uniform distribution for $\epsilon^l_{ik} \in [-1, 1]$ (circle).

rial surpasses 2^M at $M = 6$ coinciding with the magic number 7 ± 2. This could be the reason why at the magic number 7 ± 2, the separation of states collapses and chaotic dynamics takes over.

In the present Letter, we have shown that the interference between inputs drastically increases around $M \sim 7$ within the general setup of LNN. The argument of magic number 7 ± 2 presented here is only based on combinatorial arguments and does not strongly depend on the choice of parameters. Hence it is naturally expected that our explanation works for a wide class of entangled cascade networks with sigmoid units.

The term magic number 7 ± 2 was originally coined in psychology [17], where the number of chunks (items) that can be memorized in short term memory is found to be limited to about 7 ± 2. In the future, it will be interesting to search for a possible relationship with our general mechanism, noting that the magic number 7 in our case is the number of inputs beyond which the output that depends on these inputs no longer clearly separates (see also [13,18].

Note that if the coupling term ϵ_{ik} is independent of the layer, the model is a standard neural network where spin-glass theory is applicable. Numerical simulations for this case again show that chaotic dynamics are rare if M is small ($<7 \sim 10$), while they are rather common for large M (about half of the networks with randomly chosen ϵ_{ik} exhibit chaos for $M \sim 20$). However, in this case, the increase of the fraction of the networks with chaotic dynamics is not sharp. In this case, there exists stronger interlayer correlations, which suppress the folding process of phase space. In addition, the dynamic behavior largely depends on the choice of the network since single ϵ_{ik} is used recurrently. On the other hand, extension of spin-glass theory for the present layer-dependent coupling case will be a future problem.

The authors are grateful to K. Fujimoto, T. Shibata, N. Kataoka, and F. H. Willeboordse for discussion. This work was supported by a Grant-in-Aid for Scientific Research from the Ministry of Education, Science, and Culture of Japan (11CE2006).

*Electronic address: shuji@complex.c.u-tokyo.ac.jp

[1] B. D. Gomperts, I. M. Kramer, and P. E. R. Tatham, *Signal Transduction* (Academic, New York, 2002).
[2] F. Rosensenblatt, *Principle of Neurodynamics* (Spartan, Washington, D.C., 1961).
[3] W. Pitts and W. S. McCulloch, Bull. Math. Biol. **9**, 127 (1947).
[4] M. L. Minsky and S. A. Rapert, *Perceptrons* (The MIT Press, Cambridge, MA, 1988), expanded edition.
[5] J. Hertz, A. Krogh, and R. G. Palmer, *Introduction to the Theory of Neural Computation* (Addison-Wesley, Reading, MA, 1991).
[6] L. Yu, E. Ott, and Q. Chen, Phys. Rev. Lett. **65**, 2935 (1990).
[7] From a dynamical systems point of view, the above process is similar to [6] where a change of the parameter value makes the effective maximum Lyapunov exponent positive.
[8] This transition around $M \sim 7$ is observed irrespectively of β, as long as it is not too small (say larger than 2). Note that the coupling form $\beta/\sqrt{M}$ in Eq. (1) is originally adopted so that M dependence disappears asymptotically for large M, but this form is useful even for small M. Indeed the critical value of β at which the stability of the state $\vec{x} = 0$ is lost is only 15% larger than its asymptotic value 1.0, even for $M = 4$ [9].
[9] In case of $\theta^l_i = 0$, the linear stability of the state $\vec{x}_l = 0$ is lost at $\beta = \sqrt{M/2}\exp[-\psi(M/2)/2](\sim 1)$, where $\psi(x)$ is a polyGamma function [10]. For large M, $\psi(M/2) \sim \log(M/2)$ and critical β is 1.0 asymptotically.
[10] A. Crisanti, G. Paladin, and A. Vulpiani, *Products of Random Matrices* (Springer-Verlag, Berlin, 1992).
[11] This crossover around $M \sim 7$ is clearly visible as long as l is not too small. For small l (say less than 10), the tendency of the transition around $M \sim 7$ is observed but is blurred, since the dynamics are still in the transient regime.
[12] S. Ishihara and K. Kaneko (to be published).
[13] K. Kaneko, Phys. Rev. E **66**, 055201(R) (2002).
[14] K. Kaneko, Phys. Rev. Lett. **78**, 2736 (1997); Physica D (Amsterdam)**124**, 322 (1998).
[15] M. Timme, F. Wolf, and T. Geisel, Phys. Rev. Lett. **89**, 154105 (2002).
[16] The folding cannot occur in a single layer because the present system separates the *output set* linearly, as is well known. At least two layers are necessary for a folding processes.
[17] G. A. Miller, *The Psychology of Communication* (Basic Books, New York, 1975).
[18] J. S. Nicolis and I. Tsuda, Bull. Math. Biol. **47**, 343 (1985).