



**HAL**  
open science

# Biologically meaningful update rules increase the critical connectivity of generalized kauffman networks

Dominik M. Wittmann, Carsten Marr, Fabian J. Theis

## ► To cite this version:

Dominik M. Wittmann, Carsten Marr, Fabian J. Theis. Biologically meaningful update rules increase the critical connectivity of generalized kauffman networks. *Journal of Theoretical Biology*, 2010, 266 (3), pp.436. 10.1016/j.jtbi.2010.07.007 . hal-00616255

**HAL Id: hal-00616255**

**<https://hal.science/hal-00616255>**

Submitted on 21 Aug 2011

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## Author's Accepted Manuscript

Biologically meaningful update rules increase the critical connectivity of generalized kauffman networks

Dominik M. Wittmann, Carsten Marr, Fabian J. Theis

PII: S0022-5193(10)00356-5  
DOI: doi:10.1016/j.jtbi.2010.07.007  
Reference: YJTBI6069

To appear in: *Journal of Theoretical Biology*

Received date: 14 February 2010  
Revised date: 27 May 2010  
Accepted date: 12 July 2010

Cite this article as: Dominik M. Wittmann, Carsten Marr and Fabian J. Theis, Biologically meaningful update rules increase the critical connectivity of generalized kauffman networks, *Journal of Theoretical Biology*, doi:[10.1016/j.jtbi.2010.07.007](https://doi.org/10.1016/j.jtbi.2010.07.007)

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting galley proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.



[www.elsevier.com/locate/jtbi](http://www.elsevier.com/locate/jtbi)

# Biologically meaningful update rules increase the critical connectivity of generalized Kauffman networks

Dominik M. Wittmann<sup>a,b</sup>, Carsten Marr<sup>a</sup>, Fabian J. Theis<sup>a,b</sup>

<sup>a</sup>*Computational Modeling in Biology, Institute for Bioinformatics and Systems Biology, Helmholtz Zentrum München - German Research Center for Environmental Health, Ingolstädter Landstrasse 1, 85764 Neuherberg, Germany*

<sup>b</sup>*Centre for Mathematical Sciences, Technische Universität München, Boltzmannstrasse 3, 85748 Garching, Germany*

---

## Abstract

We generalize random Boolean networks by softening the hard binary discretization into multiple discrete states. These multistate networks are generic models of gene regulatory networks, where each gene is known to assume a finite number of functionally different expression levels. We analytically determine the critical connectivity that separates the biologically unfavorable frozen and chaotic regimes. This connectivity is inversely proportional to a parameter which measures the heterogeneity of the update rules. Interestingly, the latter does not necessarily increase with the mean number of discrete states per node. Still, allowing for multiple states decreases the critical connectivity as compared to random Boolean networks, and thus leads to biologically unrealistic situations.

Therefore, we study two approaches to increase the critical connectivity. First, we demonstrate that each network can be kept in its frozen regime by sufficiently biasing the update rules. Second, we restrict the randomly chosen update rules to a subclass of biologically more meaningful functions. These functions are characterized based on a thermodynamic model of gene regu-

lation. We analytically show that their usage indeed increases the critical connectivity. From a general point of view, our thermodynamic considerations link discrete and continuous models of gene regulatory networks.

*Keywords:* dynamic systems on graphs, multilevel logic, thermodynamics, transcriptional gene regulation

*2010 MSC:* 80A30, 82C20, 82C27, 92C42

---

## 1. Introduction

*Boolean networks* are a class of discrete dynamical systems. The system's  $N$  variables take only discrete values 0 or 1 and develop in discrete time steps. At each time point, the value of a variable is determined by a so-called *update rule* that deterministically depends upon the values of some of the other variables, the so-called *inputs*, at the previous time point. In 1969, Kauffman proposed random Boolean networks — so called *Kauffman networks* (KN) — as generic models for large-scale gene regulatory networks (Kauffman, 1969). Computational experiments showed that these networks exhibit surprisingly ordered structures and are able to give insights into biological phenomena such as cell replication or lineage differentiation.

Interest in KNs was rekindled as their close relation to classical models from *statistical mechanics* was realized. In a number of studies (Derrida and Pomeau, 1986; Derrida and Stauffer, 1986; Flyvbjerg, 1988) the *self-organizing capacity* of KNs was analyzed. It was shown that depending on their connectivity, KNs exhibit *ordered (frozen)* as well as *chaotic* behaviors with a *critical boundary* separating both regimes. The ordered regime is characterized by small stable attractors, whereas in the chaotic regime

long-periodic orbits frequently occur. These properties render both regimes unfavorable for the evolution of living organisms. Consequently, Kauffman promoted the idea of "living at the edge of chaos" (Kauffman, 1993). Interestingly, the critical connectivity of KNs is 2, which agrees well with the average connectivities of well-studied gene regulatory networks, e.g. in *E. coli*, *S. cerevisiae* and *B. subtilis* (Balleza et al., 2008).

Besides studies of large-scale random Boolean networks, many small- and medium-scale Boolean models of different biological processes have been manually curated and analyzed (Albert and Othmer, 2003; Li et al., 2004; Saez-Rodriguez et al., 2007; Wittmann et al., 2009a). Here however, the discretization of continuous biological quantities, such as mRNA or protein concentrations, into binary 'on'-'off' categories is often arbitrary as well as insufficient. In fact, it has been demonstrated that genes may well have more than two functionally different expression levels (Setty et al., 2003). To alleviate this issue, in some modeling applications the Boolean categories are extended to multiple discrete states, leading to *multistate logical models* (Thomas, 1991; Sánchez and Thieffry, 2001; Fauré et al., 2009).

Studies of KNs with multiple states, so-called *multistate Kauffman networks* (MKN), are scarce (Solé et al., 2000) and restricted to the biologically implausible case where all nodes have the same number of discrete states. In this case it has been observed that the critical connectivity tends to 1, as the number of states grows. Hence, it drops below the average connectivities we typically find in gene regulatory networks, cf. again Balleza et al. (2008), leading to biologically unrealistic scenarios. For (Boolean) KNs two approaches are known that can lead to an increased critical connectivity.

- (i) First, it has been shown that for any connectivity a KN can be kept in the frozen regime by biasing the update rules towards one of the discrete states (Bastolla and Parisi, 1996).
- (ii) Second, the issue of *biologically meaningful* update rules has attracted considerable attention, for a brief review see section 6.3.

In this contribution, we generalize these approaches to MKNs and ask how they affect their critical connectivity. We begin by investigating a very general class of MKNs in section 2. In particular, contrary to previous studies, the number of discrete states is not fixed for all nodes but follows a distribution. We show that the above mentioned transition between a frozen and a chaotic regime takes place in MKNs of this class. The critical boundary is analytically determined as a relation between the mean connectivity and a parameter describing the heterogeneity of the update rules. Our results are an extension of the (binary) Boolean case. They show that allowing nodes to assume multiple states lowers the critical connectivity as compared to (Boolean) KNs, the limit being 1. Interestingly however, the critical connectivity does not necessarily decrease with the mean number of states per node. With respect to our motivating question, we find in section 3 that — similar to (Boolean) KNs — each MKN can be kept in the frozen regime by biasing the update rules towards one of the discrete states: Approach (i) can be generalized to MKNs.

The main part of this manuscript is devoted to the generalization of approach (ii). In section 4 we propose a tentative characterization of biologically plausible update functions for MKNs. This characterization is motivated by

a common property of the update rules from some manually curated multistate logical models, which we call *single-switch condition*. It is further corroborated by a thermodynamic model of transcriptional gene regulation. Using this model we show that under certain assumptions regarding *cooperative effects* the promotor activity as a function of the transcription factor concentrations always satisfies the single-switch condition. From a general point of view, our thermodynamic considerations link discrete and continuous models of gene regulatory networks.

The effect of restricting update rules to the class of biologically more meaningful single-switch functions is studied in section 5. Analytic results demonstrate that the use of single-switch functions indeed increases the critical connectivity in MKNs as compared to fully random update rules. Hence, our motivating question can be answered positively also with respect to approach (ii). The usage of single-switch functions causes further interesting effects. We present an example, in which for an increasing number of states the critical connectivity increases if single-switch functions are used, while it decreases if the update rules are chosen randomly. Section 6 concludes our manuscript by a non-technical summary of our results and their discussion, in particular, in the light of previous work.

## **2. Phase transitions in multistate Kauffman networks**

In this section we introduce a very general class of MKNs and investigate it with respect to critical phenomena.

### 2.1. A general class of multistate Kauffman networks

A *discrete multistate network* consists of  $N$  variables  $x_1, x_2, \dots, x_N$  each taking values in a discrete finite set  $\Sigma_i$  with cardinality  $S_i$ , w.l.o.g.  $\Sigma_i = \{0, 1, \dots, S_i - 1\}$ ,  $i = 1, 2, \dots, N$ . Time is discretized,  $t = 0, 1, \dots$ , and we denote the variables' time courses by  $x_i(t)$ . The  $i$ -th variable is influenced by  $K_i$  inputs  $x_{i1}, x_{i2}, \dots, x_{iK_i}$  and its value at time  $t + 1$  is given by an update function  $f_i : \prod_{k=1}^{K_i} \Sigma_{ik} \rightarrow \Sigma_i$ ,

$$x_i(t + 1) = f_i(x_{i1}(t), x_{i2}(t), \dots, x_{iK_i}(t)) .$$

We call  $K_i$  the connectivity of  $x_i$ . The state of a multistate network at time  $t$  is denoted by  $X(t) = (x_1(t), x_2(t), \dots, x_N(t)) \in \prod_{i=1}^N \Sigma_i$ . A *Boolean network* is the special case of a multistate network where all  $S_i = 2$ .

We now define MKNs generalizing the definition of KNs as given e.g. in Aldana et al. (2003). A MKN is a multistate network where

- (K1) the  $K_i$  are chosen randomly from a probability distribution  $P_{\text{in}}(K)$ ,  
 $K = 1, 2, \dots, K_{\text{max}}, K_{\text{max}} \leq N$ ,
- (K2) the  $K_i$  inputs  $x_{i1}, x_{i2}, \dots, x_{iK_i}$  of  $x_i$  are chosen randomly with uniform probability from among the system's variables  $x_1, x_2, \dots, x_N$ ,
- (K3) the number of states  $S_i$  of  $x_i$  is chosen randomly from a probability distribution  $P_{\text{nos}}(S)$ ,  $S = 2, 3, \dots, S_{\text{max}}$ ,
- (K4) the values of  $f_i$  are independent and chosen randomly from a probability distribution  $P_{S_i}(s)$ ,  $s \in \Sigma_i$ .



Note that in (K4) the distribution  $P_{S_i}$  does not depend on the node  $x_i$  but only on the number of states  $S_i$  of node  $x_i$ . In particular, in the Boolean case the update rules evaluate to 0 with a certain probability  $w$  and to 1 with probability  $1 - w$ .

## 2.2. The critical boundary

From (K4) it follows that the probability  $p_{S_i}$  for the function  $f_i$  to yield two different values for two different arguments depends only on  $P_{S_i}$  and is given by

$$p_{S_i} = \sum_{s \in \Sigma_i} P_{S_i}(s) (1 - P_{S_i}(s)) . \quad (1)$$

We remark that  $p_{S_i}$  becomes maximal if  $P_{S_i}$  is the discrete uniform distribution,

$$P_{S_i}(s) = \frac{1}{S_i} , \quad s \in \Sigma_i , \quad (2)$$

cf. Appendix A. In this case,

$$p_{S_i} = \frac{S_i - 1}{S_i} ,$$

which also goes to show that  $p_{S_i} < 1$  for any choice of  $P_{S_i}$ .

Following e.g. Aldana (2003) we distinguish between the frozen and the chaotic regime by looking at the overlap  $o(t)$  between two distinct time courses  $X(t) = (x_1(t), x_2(t), \dots, x_N(t))$  and  $X'(t) = (x'_1(t), x'_2(t), \dots, x'_N(t))$  of the variables  $x_1, x_2, \dots, x_N$ . The overlap at time  $t$  is defined as the fraction of variables that take the same value in  $X(t)$  and  $X'(t)$ , i.e.

$$o(t) \stackrel{\text{def}}{=} \frac{1}{N} \sum_{i=1}^N \delta_{x_i(t), x'_i(t)} . \quad (3)$$

We now express  $o(1)$  in terms of  $o(0)$ . Fixing some  $i$ , there are two (mutually exclusive) possibilities that at  $t = 1$  we have  $x_i(1) = x'_i(1)$ :

1. All  $K_i$  inputs are equal at  $t = 0$ . This occurs with probability  $o(0)^{K_i}$ .
2. Not all inputs are equal at  $t = 0$  but  $f_i$  still yields the same value. This occurs with probability  $(1 - o(0)^{K_i}) (1 - p_{S_i})$ .

Hence, the probability for " $x_i(1) = x'_i(1)$ " is given by

$$o(0)^{K_i} + (1 - o(0)^{K_i}) (1 - p_{S_i})$$

and depends only on  $K_i$  and  $p_{S_i}$ . It follows that

$$o(1) = \sum_{K=1}^{K_{\max}} P_{\text{in}}(K) \sum_{S=2}^{S_{\max}} P_{\text{nos}}(S) [o(0)^K + (1 - o(0)^K) (1 - p_S)] .$$

We now apply a trick devised by Derrida and Pomeau (1986) called *annealed approximation*. Although the model is quenched, i.e. fixed for all time points, we assume that after each time step a new model is generated according to (K1)–(K4). It has been shown that in the *thermodynamic limit*  $N \rightarrow \infty$  the quenched and the annealed model show the same phase transition (Hilhorst and Nijmeijer, 1987). Therefore, for  $N \gg K_{\max}$  we can write

$$o(t+1) = O(o(t)) \\ \stackrel{\text{def}}{=} \sum_{K=1}^{K_{\max}} P_{\text{in}}(K) \sum_{S=2}^{S_{\max}} P_{\text{nos}}(S) [o(t)^K + (1 - o(t)^K) (1 - p_S)] \quad (4)$$

at each time step.

In the limit  $t \rightarrow \infty$  this dynamic equation becomes a fixed point equation  $o = O(o)$  for the stationary value of the overlap. Note that  $o^* = 1$  is always a fixed point. If  $o^*$  is attractive, initial differences between time courses vanish over time and the system is in an ordered regime, otherwise it is chaotic. Hence, the ordered and chaotic regimes as well as the critical boundary can

be characterized by a stability analysis of  $o^*$ . As detailed in Appendix B this yields

$$\bar{p}\bar{K} \begin{cases} < 1 & \text{ordered regime} \\ = 1 & \text{critical boundary} \\ > 1 & \text{chaotic regime ,} \end{cases} \quad (5)$$

where

$$\bar{p} \stackrel{\text{def}}{=} \sum_{S=2}^{S_{\max}} P_{\text{nos}}(S) p_S \quad (6)$$

and

$$\bar{K} \stackrel{\text{def}}{=} \sum_{K=1}^{K_{\max}} P_{\text{in}}(K) K .$$

In other words, the critical connectivity is given by

$$\bar{K}^{\text{crit}} = \frac{1}{\bar{p}} . \quad (7)$$

We observe that the phase transition depends only on the first moments  $\bar{K}$  and  $\bar{p}$ . Typically,  $\bar{K}$  can be easily computed once the underlying network structure is known. The computation of  $\bar{p}$  is more involved.

### 2.3. The situation of fully random update rules without bias

As a preparation for section 3 let us now study the multistate counterpart of the original KN as introduced by Kauffman (1969). This is to say we consider the completely unbiased situation where the  $P_{S_i}$  are the discrete uniform distributions from (2). We already remarked that in this case

$$p_{S_i} = \frac{S_i - 1}{S_i} . \quad (8)$$

Substituting (8) in (6) yields

$$\bar{p} = \sum_{S=2}^{S_{\max}} P_{\text{nos}}(S) \frac{S-1}{S} = 1 - \mu ,$$

where

$$\mu \stackrel{\text{def}}{=} \sum_{S=2}^{S_{\max}} P_{\text{nos}}(S) \frac{1}{S}.$$

Clearly, for any  $P_{\text{nos}}$ ,  $\mu$  is bounded,  $0 < \mu \leq 1/2$ . For the critical connectivity from (7)

$$\overline{K}^{\text{crit}} = \frac{1}{1 - \mu}, \quad (9)$$

this implies  $1 < \overline{K}^{\text{crit}} \leq 2$ . Moreover,  $\overline{K}^{\text{crit}}$  lowers down to 1 as  $\mu$  becomes small.

Hence, the critical connectivity in unbiased MKNs is smaller than in unbiased (Boolean) KNs. This important result serves as the motivation for our investigations in sections 3, 4 and 5. Before, let us visualize this result in two examples.

#### 2.4. Examples

Let us conclude this section by computing  $\mu$  in two special cases of  $P_{\text{nos}}$ .

- First, a delta distribution

$$P_{\text{nos}}(S) = \delta_{S, \overline{S}}, \quad S = 2, 3, \dots, S_{\max}, \quad (10)$$

for a fixed number of states  $\overline{S}$  per node. In this case,  $\mu = 1/\overline{S}$ , and from (9) we obtain the condition

$$\overline{K}^{\text{crit}} = \frac{\overline{S}}{\overline{S} - 1}, \quad (11)$$

which was already found in Solé et al. (2000). Note that in the Boolean case,  $\overline{S} = 2$ , we recover the critical connectivity 2 of KNs. For an increasing (fixed) number of states  $\overline{S}$  the critical connectivity approaches 1.

- Second, a discrete uniform distribution

$$P_{\text{nos}}(S) = \frac{1}{S_{\text{max}} - 1}, \quad S = 2, 3, \dots, S_{\text{max}}.$$

In this case,

$$\mu = \sum_{S=2}^{S_{\text{max}}} \frac{1}{S_{\text{max}} - 1} \frac{1}{S} = \frac{H_{S_{\text{max}}} - 1}{S_{\text{max}} - 1},$$

where  $H_{S_{\text{max}}}$  is the  $S_{\text{max}}$ -th harmonic number, and from (9) we obtain the criticality condition

$$\bar{K}^{\text{crit}} = \frac{S_{\text{max}} - 1}{S_{\text{max}} - H_{S_{\text{max}}}}. \quad (12)$$

It holds that  $\bar{K}^{\text{crit}} \rightarrow 1$  as  $S_{\text{max}} \rightarrow \infty$ . A better idea of this convergence is given by the well-known approximation

$$H_{S_{\text{max}}} = \ln(S_{\text{max}}) + \gamma + \mathcal{O}(S_{\text{max}}^{-1}),$$

where  $\gamma$  is the Euler-Mascheroni constant  $\gamma \approx 0.5772$ .

The critical connectivities from (11) and (12) are plotted in Figure 1A. We observe that in both examples a growing (fixed) number of states  $\bar{S}$  or a growing upper bound  $S_{\text{max}}$ , respectively, lower the critical connectivity down to 1. In the following sections we generalize the approaches (i) and (ii) from section 1 to MKNs and investigate if they lead to an increased critical connectivity.

### 3. Multistate Kauffman networks with biased update rules

In the original KN, each Boolean update function was drawn randomly from the ensemble of all Boolean functions according to the uniform distribution. In subsequent studies, however, it is mostly assumed that each function

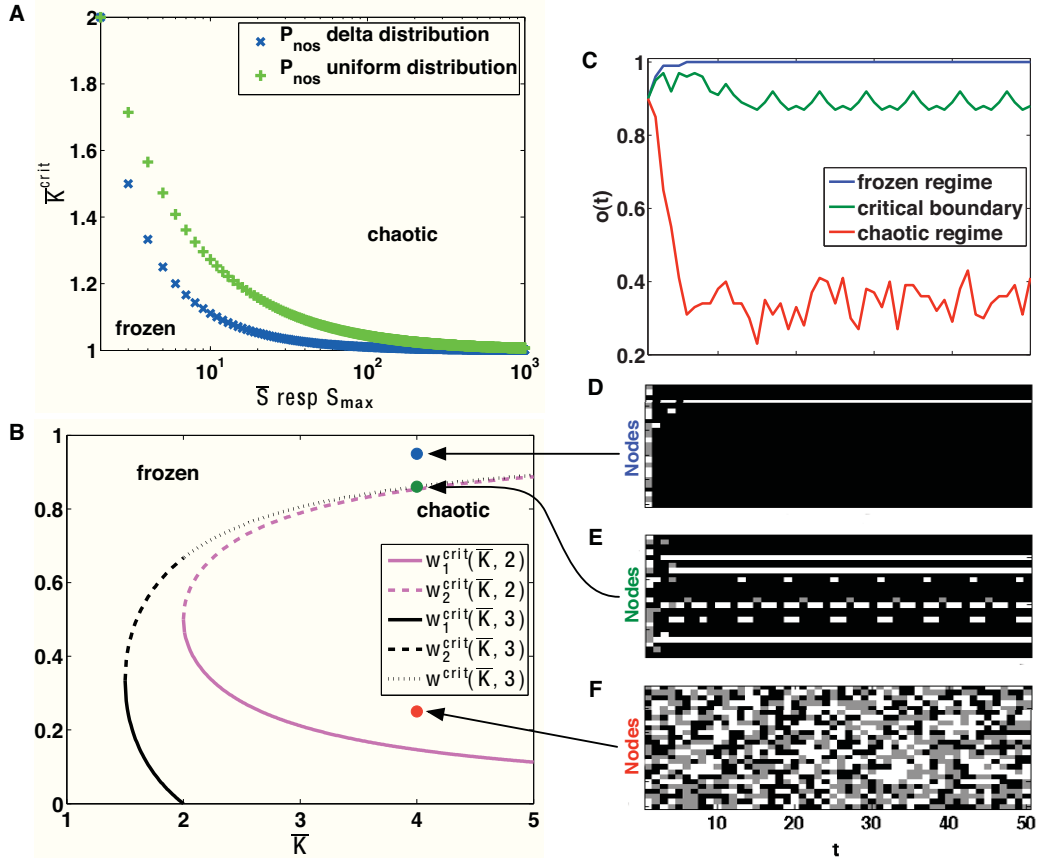


Figure 1: **(A)** The case of unbiased update rules. Critical connectivities  $\bar{K}^{\text{crit}}(\bar{S})$  from (11) (blue 'x') and  $\bar{K}^{\text{crit}}(S_{\text{max}})$  from (12) (green '+'). **(B)** The case of biased update rules. Critical boundaries  $w^{\text{crit}}(\bar{K}, \bar{S})$  from section 3.2. For  $\bar{S} = 2$  the two branches  $w_{1/2}^{\text{crit}}(\bar{K}, 2)$  from (D.1) are shown. For  $\bar{S} = 3$  the critical boundary consists of three parts:  $w_{1/2}^{\text{crit}}(\bar{K}, 3)$  from (D.2) and  $w^{\text{crit}}(\bar{K}, 3)$  from (D.3). **(C)** Plots of the overlap  $o(t)$  from (3) between two initially differing time courses for a frozen, critical and chaotic network. Each network has  $N = 100$  nodes, each node has  $\bar{S} = 3$  states and is connected to  $\bar{K} = 4$  inputs. The update rules are constructed according to distribution (13). For the frozen/critical/chaotic network we set  $w = 0.95 / w = w^{\text{crit}}(4, 3) \approx 0.8604$  from (D.3) /  $w = 1/3$ . The initial overlap was chosen to be  $o(0) = 0.9$ . **(D,E,F)** The first 50 time steps of evaluations of the frozen, critical and chaotic networks from (C); only the first 25 nodes are shown. The three states are color coded in black, grey and white. Note that subfigures C,D,E,F all have the same X-axis, namely time running from 1 to 50 as indicated in (F).

evaluates to 0 with a certain probability  $0 < w < 1$  (and to 1 with probability  $1 - w$ ). Biologically speaking, it is assumed that a gene is expressed with probability  $w$ . A natural generalization hereof to multistate systems is the distribution

$$P_{S_i}(s) = \begin{cases} w & s = 0 \\ \frac{1-w}{S_i-1} & s > 0 \end{cases} \quad (13)$$

for some  $0 < w < 1$ .

In the Boolean case, it turned out that for each connectivity a system can be kept in the frozen regime by choosing  $w$  sufficiently large or small (Bastolla and Parisi, 1996). Is this still true for MKNs?

### 3.1. The critical boundary

To answer this question, let us assume that each  $P_{S_i}$  is given by (13) for some  $0 < w < 1$ . Then (1) becomes

$$p_{S_i}^w = 2w(1-w) + (1-w)^2 \frac{S_i-2}{S_i-1},$$

and substituting this in (6) gives

$$\begin{aligned} \bar{p}^w &= \sum_{S=2}^{S_{\max}} P_{\text{nos}}(S) \left[ 2w(1-w) + (1-w)^2 \frac{S-2}{S-1} \right] \\ &= 2w(1-w) + (1-w)^2 (1 - \mu_{-1}), \end{aligned} \quad (14)$$

where

$$\mu_{-1} \stackrel{\text{def}}{=} \sum_{S=2}^{S_{\max}} P_{\text{nos}}(S) \frac{1}{S-1}.$$

We observe that  $0 < \mu_{-1} \leq 1$ , where the upper bound is exact iff  $P_{\text{nos}}(S) = \delta_{S,2}$ , i.e. in the Boolean case.

Substituting (14) in (5) gives rise to a quadratic equation in  $w$ . Solving this equation for  $w^{\text{crit}}(\bar{K}, \mu_{-1})$  yields (cf. Appendix C):

(a) For  $\bar{K} < 1 + \mu_{-1}$  the system is always frozen.

(b) For  $\bar{K} = 1 + \mu_{-1}$  the system is critical if and only if

$$w = w^{\text{crit}}(1 + \mu_{-1}, \mu_{-1}) = \frac{\mu_{-1}}{\mu_{-1} + 1},$$

otherwise it is frozen.

(c) In the **Boolean case**, i.e.  $\mu_{-1} = 1$ , and for  $\bar{K} > 1 + \mu_{-1} = 2$  the critical boundary is described by

$$w_{\frac{1}{2}}^{\text{crit}}(\bar{K}, 1) = \frac{1}{2} \left[ 1 \mp \sqrt{1 - \frac{2}{\bar{K}}} \right] \in (0, 1).$$

For

$$w_1^{\text{crit}}(\bar{K}, 1) < w < w_2^{\text{crit}}(\bar{K}, 1)$$

the system is chaotic, otherwise it is frozen. This agrees with previous results about Boolean KNs, see e.g. Bastolla and Parisi (1996).

(d) For  $\mu_{-1} < 1$  and  $1/(1 - \mu_{-1}) > \bar{K} > 1 + \mu_{-1}$  the critical boundary is described by

$$w_{\frac{1}{2}}^{\text{crit}}(\bar{K}, \mu_{-1}) = \frac{1}{\mu_{-1} + 1} \left( \mu_{-1} \mp \sqrt{1 - \frac{\mu_{-1} + 1}{\bar{K}}} \right) \in (0, 1).$$

For

$$w_1^{\text{crit}}(\bar{K}, \mu_{-1}) < w < w_2^{\text{crit}}(\bar{K}, \mu_{-1})$$

the system is chaotic, otherwise it is frozen. For  $\bar{K} = 1/(1 - \mu_{-1})$ ,  $w_1^{\text{crit}} = 0$  is not a valid solution for  $w$  anymore, as we require  $w \in (0, 1)$ .



(e) For  $\mu_{-1} < 1$  and  $\bar{\mathbf{K}} \geq 1/(1 - \mu_{-1})$  the critical boundary is described by

$$w^{\text{crit}}(\bar{K}, \mu_{-1}) = \frac{1}{\mu_{-1} + 1} \left( \mu_{-1} + \sqrt{1 - \frac{\mu_{-1} + 1}{\bar{K}}} \right) \in (0, 1),$$

for larger values of  $w$  the system is frozen, otherwise it is chaotic. It can easily be seen that  $w^{\text{crit}}(\bar{K}, \mu_{-1})$  is monotonous in both its arguments and that  $w^{\text{crit}} \rightarrow 1$  as  $\bar{K} \rightarrow \infty$ .

We chose to solve the criticality condition for  $w$  (and not  $\bar{K}^{\text{crit}}$ ), as from this presentation we easily see that for any values of  $\bar{K}$  and  $\mu_{-1}$  a system is kept in its frozen regime by choosing a sufficiently large  $w$ . Thus, our motivating question can be answered in the affirmative: Biasing update rules towards one of the discrete states indeed increases the critical connectivity, from a theoretical point of view, even beyond any bound.

### 3.2. Examples and visualization

For clarification and illustration purposes, let us again consider the example where  $P_{\text{nos}}$  follows the delta distribution from (10) for a fixed number of states  $\bar{S}$ . This entails  $\mu_{-1} = 1/(\bar{S} - 1)$ , and we can express the classification from section 3.1 in terms of  $\bar{S}$ , cf. Appendix D. This classification is visualized in Figure 1B. We observe that  $w^{\text{crit}}(\bar{K}, \bar{S})$  is monotonously increasing in  $\bar{S}$  and  $\bar{K}$ , and that  $w^{\text{crit}} \rightarrow 1$  as  $\bar{K} \rightarrow \infty$ .

To further visualize our results, we generated a frozen, critical and chaotic MKN, for details see caption of Figure 1C. This figure shows the overlap  $o(t)$  from (3) between two initially differing time courses. For each network one of these time courses is plotted in Figures 1D–F. In the frozen network the

overlap quickly reaches its maximal value 1 indicating that both time courses have reached the steady state shown in Figure 1D. While the critical network exhibits clear short-periodic oscillations after approximately 10 time steps, cf. Figure 1E, no pattern or periodicity is discernible in the chaotic network in Figure 1F. (We know, of course, that ultimately also the chaotic network will fall into a limit cycle due to the finite size of the state space.) If we discard the first 10 time steps, the overlap in the critical network oscillates around 0.88 and fluctuates around 0.34 in the chaotic network. The latter ratio agrees well with the expected overlap of two randomly drawn configurations.

#### 4. A class of biologically meaningful update rules

In the last sections, the update functions  $f_i$  were chosen from the set of all possible update rules according to (K4). However, as argued below, especially for higher number of states most of these update rules are arguably not plausible from a biological point of view. In this section, we characterize a class of biologically more meaningful update rules for MKNs. We illustrate the main characteristics of this class in section 4.1. Subsequently, in section 4.2, we present a more formal derivation using a thermodynamic model of gene regulation. MKNs with biologically meaningful update rules are then investigated in section 5, especially with respect to their critical connectivity.

##### 4.1. Update rules with single-switch inputs

Figure 2A shows an example of a random multistate update rule. Is such a function really plausible from a biological point of view? To answer this question let us take a look at the manually curated update rules from multistate logical models, e.g. of the yeast cell cycle (Fauré et al., 2009) or

the *Drosophila* Gap-gene System (Sánchez and Thieffry, 2001). The "least common denominator" of these rules is, arguably, that there is only one threshold for each input at which the function changes its output. In other words, there is only a single switching point along each input dimension. We will call functions satisfying this condition *single-switch functions*. More formally, a single-switch function  $f$  with inputs  $x_1 \in \Sigma_1, x_2 \in \Sigma_2, \dots, x_K \in \Sigma_K$  has the property that for each  $1 \leq k \leq K$  there exists a threshold  $\theta_k \in \Sigma_k$  such that  $\text{sgn}(x_k - \theta_k) = \text{sgn}(x'_k - \theta_k)$  implies  $f(x_1, \dots, x_k, \dots, x_K) = f(x_1, \dots, x'_k, \dots, x_K)$ . An example of a single-switch function is shown in Figure 2B.

One might wonder why we allow multiple states for the input variables if in the end they behave essentially like Boolean (binary) variables. This is certainly a valid argument if we consider only one update rule. Typically however, a transcription factor influences more than one gene and the thresholds for these regulations can differ. This is, in fact, precisely the reasoning that led to the development of multistate logical models. We would like to refer the reader to the review by Thomas (1991) where the author addresses inter alia the question "When and How Should One Use Variables with More Than Two Values?"

We argue that imposing the single-switch condition is a reasonable restriction of update rules and that MKNs with single-switch update rules are closer to biological reality than MKNs with fully random update rules, see section 6.3 for further discussion. Therefore, we believe it worthwhile to investigate MKNs with single-switch update rules and we will do so in section 5. Before, let us further and more formally validate the single-switch

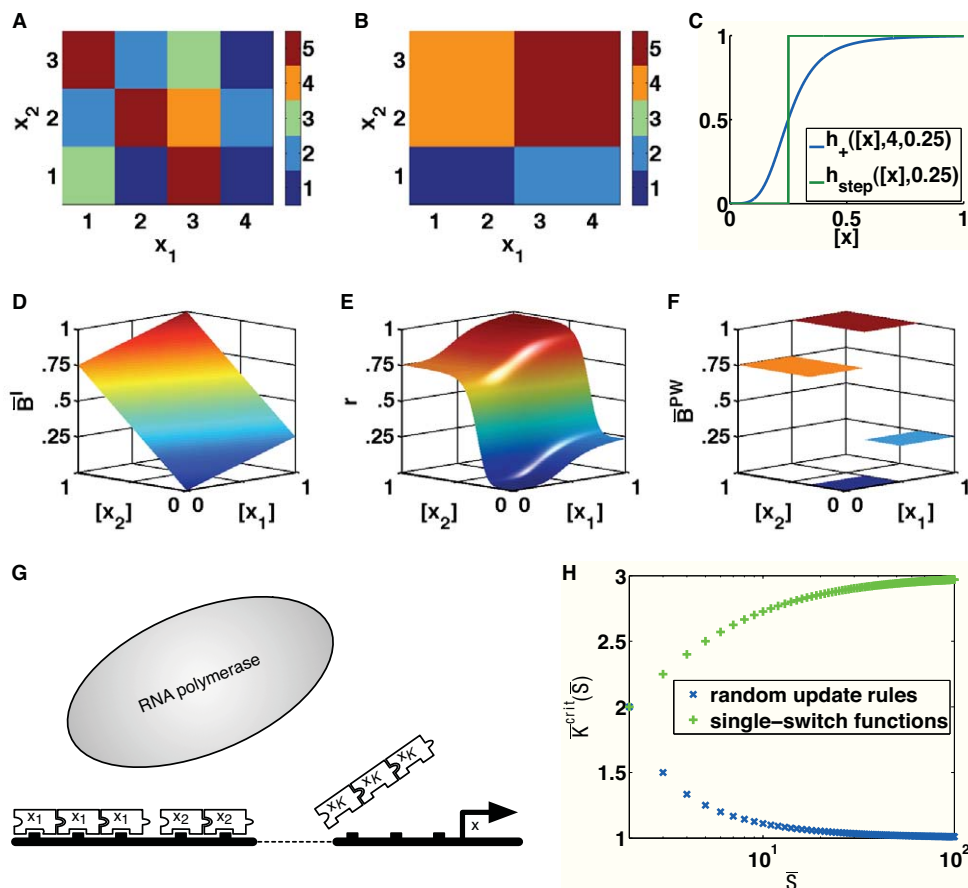


Figure 2: **A** Random update rule depending on two inputs,  $x_1 \in \{1, 2, 3, 4\}$  and  $x_2 \in \{1, 2, 3\}$ . The function values were drawn randomly from  $\{1, 2, 3, 4, 5\}$ . **B** Single-switch function depending on two inputs,  $x_1 \in \{1, 2, 3, 4\}$  and  $x_2 \in \{1, 2, 3\}$ . The function possesses one switching point along the  $x_1$ -dimension (between 2 and 3) and one switching point along the  $x_2$ -dimension (between 1 and 2). **C** Positive Hill function and its approximation by a Heaviside step function. **D, E, F** Derivation of the single-switch function shown in (B) from transcription rates  $r_{(0,0)} = 0, r_{(1,0)} = 0.25, r_{(0,1)} = 0.75, r_{(1,1)} = 1$  as discussed in section 4.4. From left to right:  $\bar{B}_{(r,\omega)}^I$  from (16), overall transcription rate  $r$  from (17) and  $\bar{B}_{(r,\omega)}^{PW}$  from (19). **G** Schematic representation of transcriptional gene regulation. Clusters of identical binding sites allow for the homotypically cooperative binding of the transcription factors  $x_1, x_2, \dots, x_K$  to the promoter region of gene  $x$ . Heterotypic cooperativity is not taken into account. **H** Critical connectivities  $\bar{K}^{\text{crit}}$  for random and single-switch update functions in terms of a fixed number of states  $\bar{S}$ . Blue markers ('x') show the critical connectivity from (11) for fully random update rules; green markers ('+') show the critical connectivity from (22) for single-switch functions.

criterion.

#### 4.2. A thermodynamic model of gene regulation

In order to provide a solid theoretical foundation for our characterization of meaningful update rules, we now describe a thermodynamic model of gene regulation, leaning heavily on Sneppen and Zocchi (2005). Here, we present only the most important results and formulas. More detailed computations and derivations can be found in Appendix E.

Let us consider a prototypic gene  $x$  that is regulated by  $K$  regulators  $x_1, x_2, \dots, x_K$ . For each regulator  $x_k$ ,  $1 \leq k \leq K$ , there are  $n_k$  binding sites in the promotor region of  $x$ , cf. Figure 2G. We assume maximal *homotypic cooperativity*, which is to say, we assume that each transcription factor  $x_k$ ,  $1 \leq k \leq K$ , can bind only as a polymer consisting of  $n_k$  monomers. Moreover, we neglect *heterotypic cooperativity*, i.e. cooperative effects between two different transcription factors. For further discussion of these assumptions see section 6.4. We thus have  $2^K$  accessible promotor configurations, which we denote by vectors  $\omega \in \{0, 1\}^K$ , the  $k$ -th component  $\omega(k)$  indicating whether or not the  $k$ -th regulator polymer is bound.

As detailed in Appendix E we can express the probability of a certain promotor configuration  $\omega$  as

$$\pi_\omega = \prod_{k|\omega(k)=1} h_+([x_k], n_k, A_k) \cdot \prod_{k|\omega(k)=0} h_-([x_k], n_k, A_k), \quad (15)$$

where

$$h_+([x_k], n_k, A_k) = \frac{A_k [x_k]^{n_k}}{1 + A_k [x_k]^{n_k}}$$

and

$$h_-([x_k], n_k, A_k) = \frac{1}{1 + A_k [x_k]^{n_k}}$$

are positive and negative Hill functions, respectively. Brackets indicate ligand concentrations, and  $A_k$  is the *association constant* describing the bonding affinity between the  $k$ -th regulator and its binding site.

We denote the transcription rate in each promotor configuration  $\omega$  by  $r_\omega$  and let  $\overline{B}_{(r_\omega)}^I$  be the  $K$ -dimensional multilinear interpolation of the points  $(\omega, r_\omega) \in \mathbb{R}^{K+1}$ ,

$$\overline{B}_{(r_\omega)}^I(y_1, y_2, \dots, y_K) = \sum_{\omega} r_\omega \prod_{k|\omega(k)=1} y_k \cdot \prod_{k|\omega(k)=0} (1 - y_k). \quad (16)$$

Using (15) and recalling that  $h_-([x_k], n_k, A_k) = 1 - h_+([x_k], n_k, A_k)$ , we can express the overall transcription rate  $r$  as

$$\begin{aligned} r &= \sum_{\omega} \pi_{\omega} r_{\omega} \\ &= \overline{B}_{(r_\omega)}^I(h_+([x_1], n_1, A_1), h_+([x_2], n_2, A_2), \dots, h_+([x_K], n_K, A_K)) , \quad (17) \end{aligned}$$

i.e. as the composition of the multilinear interpolation with Hill functions. Hence, we can nicely separate linearities and non-linearities. Non-linearities appear only as single-component non-linearities and take the form of standard Hill functions. The different inputs are multilinearly coupled.

In discrete models of gene regulation the overall transcription rate  $r$  of gene  $x$  as a (continuous) function of the transcription factor concentrations  $[x_k]$ ,  $1 \leq k \leq K$ , is mimicked by a discrete update rule. This process is formalized in the next section. It will turn out that update rules obtained from transcription rates  $r$  as in (17) by such a discretization are always single-switch functions.

### 4.3. The transition from continuous to discrete models of gene regulation

As a first step towards a discrete model we replace the sigmoidal Hill functions  $h_+([x_k], n_k, A_k)$  in (17) by Heaviside step functions

$$h_{\text{step}}([x_k], D_k^{\text{micro}}) = \begin{cases} 0, & [x_k] < D_k^{\text{micro}} \\ 1, & [x_k] \geq D_k^{\text{micro}} \end{cases}, \quad (18)$$

where  $D_k^{\text{micro}} = \sqrt[n_k]{1/A_k}$  is the so-called *microscopic dissociation constant*, which indicates the concentration level for half-maximal activation, cf. Figure 2C. This results in *piecewise linear* functions

$$\bar{B}_{(r_\omega)}^{PW} = \bar{B}_{(r_\omega)}^I (h_{\text{step}}([x_1], D_1^{\text{micro}}), \dots, h_{\text{step}}([x_K], D_K^{\text{micro}})) \quad (19)$$

as introduced by Glass and Kauffman (1973).

We can now easily switch to discrete variables  $x \in \Sigma = \{0, 1, \dots, S-1\}$  and  $x_k \in \Sigma_k = \{0, 1, \dots, S_k-1\}$  by choosing transcription rates  $r_\omega \in \Sigma$  and defining our update function  $f$  of  $x$  via

$$\begin{aligned} f : \prod_{k=1}^K \Sigma_k &\longrightarrow \Sigma \\ (x_1, x_2, \dots, x_K) &\longrightarrow \bar{B}_{(r_\omega)}^{PW}(x_1, x_2, \dots, x_K) . \end{aligned} \quad (20)$$

In the discrete case we can w.l.o.g. choose the thresholds  $D_k^{\text{micro}}$  from  $\Sigma_k$ . Further note that in definition (18) the threshold value itself is contained in the preimage of 1. Hence, in the discrete case, the set of non-degenerate thresholds is  $\Sigma_k \setminus \{0\}$ , as for these values there is a non-empty left- as well as right-hand side.

The resulting  $f$  from (20) is clearly a single-switch function (with switching point  $D_k^{\text{micro}}$  along the  $x_k$ -dimension) as defined above in section 4.1.

Thus, we could corroborate our characterization of meaningful update rules. In the following, we will consider update rules satisfying the single-switch condition. In particular, we ask if their usage leads to an increased critical connectivity in MKNs. Before, let us further illustrate our derivation of single-switch functions using a simple example.

#### 4.4. Example

Let us show how the single-switch function shown in Figure 2B could have arisen from the thermodynamic model described above. Assume that a gene  $x$  is regulated by two transcription factors  $x_1$  and  $x_2$ , which can bind to the promotor region of  $x$  as  $n_1 = 3$ -mers and  $n_2 = 4$ -mers, respectively, (maximal homotypic cooperativity). The bonding affinity between  $x_1$  ( $x_2$ ) and the promotor region determines the threshold  $D_1^{\text{micro}}$  ( $D_2^{\text{micro}}$ ) for the regulation of  $x$  by  $x_1$  ( $x_2$ ). Here, we take  $D_1^{\text{micro}} = 1/3$  and  $D_2^{\text{micro}} = 1/2$  in unit-normalized concentrations. We denote the possible promotor configurations by  $(0, 0)$ ,  $(1, 0)$ ,  $(0, 1)$  and  $(1, 1)$ , the first (second) component indicating whether or not  $x_1$  ( $x_2$ ) is bound. We choose (unit normalized) transcription rates  $r_{(0,0)} = 0$ ,  $r_{(1,0)} = 0.25$ ,  $r_{(0,1)} = 0.75$ ,  $r_{(1,1)} = 1$  for these different configurations.

Figure 2D shows the transcription rates  $r_\omega$ ,  $\omega \in \{0, 1\}^2$ , and their multilinear interpolation  $\bar{B}_{(r_\omega)}^I$  from (16). If we couple  $\bar{B}_{(r_\omega)}^I$  with Hill functions (with thresholds  $D_1^{\text{micro}}$ ,  $D_2^{\text{micro}}$  and exponents  $n_1$ ,  $n_2$ ) we obtain the overall transcription rate  $r$  from (17) shown in Figure 2E. Finally, replacing the Hill functions by Heaviside step functions yields the piecewise linear function  $\bar{B}_{(r_\omega)}^{PW}$  from (19) shown in Figure 2F.

Now, let us say that  $x_1$  regulates two other genes, one at a higher and one



at a lower threshold than  $D_1^{\text{micro}}$ . In this case, we would allow 4 discrete states for  $x_1$  separated by these three thresholds and  $D_1^{\text{micro}}$  would lie between the second and the third discrete state. Similarly, let  $x_2$  regulate one other gene at a higher threshold than  $D_2^{\text{micro}}$ . This would require three discrete states for  $x_2$  and  $D_2^{\text{micro}}$  would lie between the first and the second. Finally, assume that gene  $x$  is also a transcription factor and regulates four genes at different thresholds. This leads to 5 discrete states for  $x$ . After this discretization, the piecewise linear function from Figure 2F becomes the single-switch function shown in Figure 2B.

## 5. Multistate Kauffman networks with single-switch functions

In the following, we no longer draw the update rules randomly from the set of all possible rules, but modify (K4) to

(K4') Each  $f_i$  is a single-switch function with non-degenerate thresholds randomly chosen from  $\Sigma_{ik} \setminus \{0\}$  and output values (transcription rates) chosen according to the distribution  $P_{S_i}$ .

Does this increase the critical connectivity as compared to (7)? To answer this question let us analytically describe the critical boundary.

### 5.1. The critical boundary

Due to the restricted architecture of the update rules, we have to slightly modify our reasoning from section 2.2. We begin by deriving an appropriate iteration for  $o(t)$  from (3), which was defined as the overlap between two time courses  $X(t)$  and  $X'(t)$ . To this end, let us consider the  $i$ -th node,  $1 \leq i \leq N$ , and its  $k$ -th input,  $1 \leq k \leq K_i$ . We denote the probability that for two

different values  $x_{ik} \neq x'_{ik}$  of this input we have  $h_{\text{step}}(x_{ik}, \theta) \neq h_{\text{step}}(x'_{ik}, \theta)$ , where  $\theta \in \Sigma_{ik} \setminus \{0\}$  is a randomly chosen threshold, by  $q_{S_{ik}}$ . For a way to compute  $q_{S_{ik}}$ , see Appendix F. Let

$$\bar{q} \stackrel{\text{def}}{=} \sum_{S=2}^{S_{\max}} P_{\text{nos}}(S) q_S .$$

Clearly,  $\bar{q} = 1$  iff for all  $S$  with  $P_{\text{nos}}(S) \neq 0$  the support of  $P_S$  consists of exactly two points, i.e. iff the network is essentially a Boolean network.

Analogously to section 2.2, we reason that, given  $o(t)$ , there are two (mutually exclusive) possibilities that at  $t + 1$  we have  $x_i(t + 1) = x'_i(t + 1)$ .

1. For each of the  $K_i$  inputs either of the following is true:
  - 1.1. The input is contained in the overlap at  $t$ , i.e.  $x_{ik}(t) = x'_{ik}(t)$ . This occurs with probability  $o(t)$ .
  - 1.2. The input is not contained in the overlap,  $x_{ik}(t) \neq x'_{ik}(t)$ , but still  $h_{\text{step}}(x_{ik}(t), \theta) = h_{\text{step}}(x'_{ik}(t), \theta)$ . This occurs with probability  $(1 - o(t))(1 - q_{S_{ik}})$ .

The overall probability for possibility 1 is given by

$$\left[ \sum_{S=2}^{S_{\max}} P_{\text{nos}}(S) [o(t) + (1 - o(t))(1 - q_S)] \right]^{K_i} = [o(t) + (1 - o(t))(1 - \bar{q})]^{K_i} .$$

2. Possibility 1 is wrong, but  $f_i$  still yields the same value. Using (1), this occurs with probability  $\left(1 - [o(t) + (1 - o(t))(1 - \bar{q})]^{K_i}\right) (1 - p_{S_i})$ .

We can now define an iteration for the overlap  $o(t)$ .

$$o(t+1) = O^{\text{switch}}(o(t)) \\ \stackrel{\text{def}}{=} \sum_{K=1}^{K_{\max}} P_{\text{in}}(K) \sum_{S=2}^{S_{\max}} P_{\text{nos}}(S) \left[ [o(t) + (1-o(t))(1-\bar{q})]^K + \right. \\ \left. (1 - [o(t) + (1-o(t))(1-\bar{q})]^K) (1-p_S) \right].$$

Clearly,  $o^* = 1$  is a fixed point. A stability analysis of this fixed point very similar to Appendix B yields the following phase transition:

$$\bar{p}\bar{K}\bar{q} \begin{cases} < 1 & \text{ordered regime} \\ = 1 & \text{critical boundary} \\ > 1 & \text{chaotic regime.} \end{cases}$$

Consequently, the critical connectivity is given by

$$\bar{K}^{\text{crit}} = \frac{1}{\bar{p}\bar{q}}. \quad (21)$$

Comparing (21) to (7) we observe that the restriction of update rules to single-switch functions introduces the factor  $1/\bar{q}$  into the criticality condition. Since  $\bar{q} < 1$  in the non-Boolean case, this leads to an increased critical connectivity for MKNs, which answers our initiatory question.

### 5.2. Example

Let us conclude by considering the special case in which  $P_{\text{nos}}$  is the delta distribution from (10) for a fixed number of states  $\bar{S}$ , and  $P_{\bar{S}}$  is the discrete uniform distribution from (2). The (mean) probability  $\bar{q} = q_{\bar{S}}$  then computes to

$$\bar{q} = \frac{1\bar{S} + 1}{3\bar{S} - 1},$$

cf. Appendix F.

Dividing the critical connectivity from (11) by  $\bar{q}$  yields

$$\bar{K}^{\text{crit}} = 3 \frac{\bar{S}}{\bar{S} + 1}. \quad (22)$$

In Figure 2H the critical connectivity from (22) and — for comparison — the critical connectivity from (11) are shown. For larger values of  $\bar{S}$  the use of single-switch functions leads to a threefold increase in the critical connectivity as compared to fully random update rules. Interestingly, while for fully random update rules the critical connectivity decreases as  $\bar{S} \rightarrow \infty$ , it increases if single-switch functions are used.

## 6. Summary and discussion

KNs as generic models of gene regulatory networks are well studied and able to explain certain aspects of biological processes such as cell replication or lineage differentiation. Still, in many respects these models are arguably too crude a simplification. In this contribution, we proposed modifications of the original KN in order to enhance its biological plausibility.

### 6.1. Summary

MKNs can be seen as a natural generalization of (Boolean) KNs and allow us to soften the hard binary discretization. We investigated a very general class of MKNs and demonstrated that they exhibit a phase transition from frozen to chaotic behavior. In the frozen regime, networks are robust in the sense that they tolerate perturbations. In the chaotic regime, even small perturbations will lead to the divergence of trajectories and thus to qualitatively different behaviors. The critical boundary between both regimes was

determined analytically. In its most general representation (7) the critical connectivity is inversely proportional to the mean heterogeneity parameter  $\bar{p}$  from (6) of the update rules. This parameter becomes maximal if the update rules are unbiased, i.e. assume each value in the state space with equal probability. It is minimal (zero), if the update rules are constant functions.

A nice intuitive interpretation of the phase transition (5) is given in terms of "damage spreading". Assume that at time  $t$  the state of a MKN is damaged in one node, i.e. the state of this node is altered. On average this node affects  $\bar{K}$  other nodes at the next time step  $t + 1$ . The parameter  $\bar{p}$  is the mean probability that a change of input indeed leads to a change of output of an update rule. The product  $\bar{K}\bar{p}$  thus gives the mean number of damaged (changed) nodes at time  $t + 1$ . In the case  $\bar{K}\bar{p} > 1$  damage will spread through the network, in the case  $\bar{K}\bar{p} < 1$  the network is able to "repair" damage over time.

Our results show that MKNs have smaller critical connectivities than the original (Boolean) KNs. Yet we also demonstrated that each system can be kept in its frozen regime by putting a sufficiently heavy bias on one of the states, cf. Figure 1B. In the Boolean case, the critical weights  $w_{1/2}^{\text{crit}}$  are, of course, symmetric about 1/2. In MKNs there still is a range of connectivities where we can freeze a network by choosing either sufficiently large or sufficiently small weights. For higher connectivities, however, the option of choosing small weights ceases to exist. Intuitively speaking, even if we set the weight to zero, the heterogeneity among the remaining states would still be too large. Mathematically, this is reflected in the solution  $w_1^{\text{crit}}$  becoming negative. From a biological point of view, this may indicate that

in real genetic networks the update functions have a base level of activation and deviations thereof constitute well-defined exceptions.

In a next step, we restricted the randomly chosen update rules to a subclass of biologically more meaningful functions, so-called single-switch functions. These were characterized based on a thermodynamic model of gene regulation. The crucial property of single-switch functions is that their output value changes only once along each input dimension. Hence, there is a chance that a single-switch function does not "realize" a change in one of its inputs, simply because the input remains on the same side of its switching point. In terms of "damage spreading" this implies that damage is less likely to spread. (We remark that this is precisely what is measured by the additional factor  $1/\bar{q}$  in (21).) Consequently, we expect MKNs with single-switch update rules to be more robust. Indeed we found that the use of single-switch functions increases the critical connectivity; in a specific example by a factor of up to three. Similar to the study of nested canalizing functions (Kauffman et al., 2004), our results show the importance of the update functions' architecture for the dynamic behavior of KNs.

### 6.2. Previous work on multistate Kauffman networks

When employing multistate logical models one needs to make a fundamental decision about how update rules act on variables. Either the value of an update rule determines the *absolute* value of a variable at the next time step or it determines the change of the variable *relative* to the previous state, i.e. whether the variable is assigned the next higher or the next lower state. KNs with multiple states using the latter update policy are called *random walk networks*. Analyses of these networks revealed a phase transition be-

tween a chaotic and an ordered regime similar to the standard KN (Luque and Ballesteros, 2004; Ballesteros and Luque, 2005). In the chaotic regime, variables follow random-walk like trajectories, which gives the name to these networks. In here, we did not consider random walk networks but KNs with multiple states using the first update policy; we referred to them as MKNs.

Previous studies (Solé et al., 2000) of MKNs are restricted to the biologically unreasonable case in which all nodes have the same number of states. In agreement with our more general results, it has been shown that in this case an increasing number of states decreases the critical connectivity down to 1. It is tempting to assume that this stays true (at least in a qualitative sense) when replacing the (fixed) number of states by the mean of a non-degenerate distribution. However, our results demonstrate that, in general, this is wrong. From the critical connectivity in (9), for instance, we deduce that here the crucial parameter is not the mean number of states, but the mean of the reciprocal number of states  $\mu$ .

Generally speaking, the mean of a strictly positive, non-constant random variable  $R$  and the mean of  $1/R$  are related by

$$\mathbb{E}\left(\frac{1}{R}\right) > \frac{1}{\mathbb{E}(R)},$$

for a proof see e.g. Kendall et al. (1987). Furthermore, we can easily come up with a distribution

$$P(R = 2) = \frac{r-1}{r} \quad \text{and} \quad P(R = r^2) = 1/r$$

for  $r \in \mathbb{N}$ , such that  $\mathbb{E}(R) \rightarrow \infty$  but  $\mathbb{E}(1/R) \rightarrow 1/2$  as  $r \rightarrow \infty$ . Setting  $P_{\text{nos}} = P$ , the critical connectivity from (9) increases up to 2 as  $r \rightarrow \infty$ , although the mean number of states grows beyond any bound.

### 6.3. *Biologically meaningful update functions*

The issue of biologically meaningful update rules for Boolean networks has attracted considerable attention. In Harris et al. (2002) it has been observed that in lower organisms most genes have one so-called *canalyzing* (forcing) input, such that for one value of this input, the output value is fixed. Only if this input is not canalyzing, do the other inputs become relevant. In Kauffman et al. (2003) the concept of canalyzing functions was extended to nested canalyzing functions. This was done mostly for technical reasons as for nested canalyzing rules simple sampling schemes exist. In a subsequent study, networks with nested canalyzing rules were shown to be always stable (Kauffman et al., 2004).

A different class of biologically plausible Boolean rules was defined in Raeymaekers (2002) based on the assumption that each transcription factor is either an overall activator or an overall inhibitor. If one furthermore assumes that the activatory and inhibitory regulators of a gene interact in an additive fashion, one can model genetic networks by so-called *random threshold networks* (Rohlf and Bornholdt, 2002; Szejka et al., 2008). Here, the update function of a node is a weighted sum over its inputs (with positive and negative weights depending on whether the input is an activator or inhibitor), which is thresholded to yield a binary output.

Somewhat contrary to the previously mentioned studies, in Buchler et al. (2003) mechanistic models of transcriptional control show that virtually any regulatory logic function can be implemented by cis-regulatory constructs, including e.g. XOR gates. Being neither canalyzing nor monotonous, these logic functions were deemed biologically meaningless in the studies mentioned



above. In short, there is no generally accepted definition of "biologically meaningful".

Naturally, this matter becomes even more difficult, if nodes assume more than two discrete states. For this reason, we did not attempt to define *the* subset of biologically meaningful update rules. Rather, we delineated a characteristic feature — the single-switch condition — which is shared by (most) biologically meaningful update rules. We did so by analyzing manually curated multistate logical models as well as by studying a thermodynamic model of gene regulation. We do not claim that all single-switch functions are implemented in real gene regulatory networks. In other words, we deem our characterization of meaningful update rules rather conservative in the sense that the class of single-switch functions is likely too large, i.e. still contains biologically implausible rules. (Naturally, we cannot fully exclude the possibility that single update rules of real gene regulatory networks do not satisfy the single-switch condition, either.) This, of course, raises the question of how to define and impose further constraints.

#### 6.4. *A thermodynamic model of gene regulation*

From a more general point of view, the presented thermodynamic model provides the biophysical foundation for the relation between discrete and continuous models of gene regulatory networks. Our modeling approach can be nicely related to previous studies:

- In the mathematical framework for gene regulatory networks described in Mestl et al. (1995) the probabilities  $\pi_\omega$  from (15) are called indicator functions; however, no derivation of these functions is presented.

- In the Boolean case, we can think of the discretized transcription rates  $r_\omega \in \{0, 1\}$ ,  $\omega \in \{0, 1\}^K$ , as a  $K$ -variate Boolean function and the multilinear interpolation  $\overline{B}_{(r_\omega)}^I$  from (16) is equal to the Zhegalkin representation of this Boolean function (Faisal et al., 2008).
- The functions  $\overline{B}_{(r_\omega)}^{PW}$  from (19) give rise to an intermediate between discrete and continuous models: systems of piecewise linear differential equations as studied e.g. in Glass and Kauffman (1973); Edwards (2000); de Jong et al. (2004).
- The established connection between continuous and discrete models of gene regulatory networks has already been used for the standardized continuous extension of Boolean models (Wittmann et al., 2009a,b).

The main reason for the complexity of transcriptional gene regulation are cooperative effects between transcription factors. Cooperativity exists, if a bound protein increases (positive cooperativity) or decreases (negative cooperativity) the affinity of the promotor for further proteins of the same kind (homotypic cooperativity) or of a different kind (heterotypic cooperativity).

In the derivation of formula (17) we made two simplifying assumptions regarding cooperative effects: We assumed strong homotypic cooperativity and neglected heterotypic cooperativity. Strong homotypic cooperativity allows us to assume a quasi-simultaneous binding of all ligands of the same kind. Intuitively speaking, we argue that once a binding site is occupied strong interactions between the ligands lead to a fast occupation of the remaining sites. Consequently, we can discard the configurations in which the binding sites are only partially occupied. This, in turn, greatly enhances the

amenability of the resulting formulas — an important issue in 1910, when Hill conducted his famous studies (Hill, 1910).

Although nowadays computational intractability is no longer an issue, the Hill equation is still widely used due to its simplicity and variability. In fact, one can account for sequential ligand binding by allowing non-integer exponents. For example, in the case of hemoglobin, in which four oxygen molecules are known to bind with a high degree of positive cooperativity, the measured Hill coefficient ranges from 1.7 to 3.2 rather than 4 (Hill, 1910). For a comprehensive review of the limitations of the Hill equation and possible more realistic extensions, see Weiss (1997). The multivariate case of these extensions, however, is still an open issue.

With respect to our second assumption, that of negligible heterotypic cooperativity, we remark that, for example, bioinformatics analyses of the *Drosophila* genome indeed revealed significant short-range homotypic clustering of binding sites but no systematic heterotypic clustering between binding sites of different factors (Segal et al., 2008). Still, heterotypic cooperativity exists and we should be wary of this assumption. Luckily, we can argue that taking into account heterotypic cooperativity does not change the switch-like character of the overall transcription rate  $r$ . To see this, we fix all the regulator concentrations but one, w.l.o.g.  $[x_1]$ . Then we can write the overall transcription rate  $r$  in terms of  $[x_1]$

$$r = \frac{c_1 + c_2 [x_1]^{n_1}}{c_3 + c_4 [x_1]^{n_1}},$$

with constants  $c_1, c_2, c_3, c_4$ . This still is a sigmoidal interpolation between the transcription rate  $r = c_1/c_3$  for  $[x_1] = 0$  and the transcription rate  $r \approx c_2/c_4$  for large  $[x_1]$ . However, we can no longer express  $r$  in a form similar to

(17), and also the determination of the switching thresholds becomes more involved.

### 6.5. Outlook

In continuation of our work, several further unrealistic features of KNs could be addressed. First, the synchronous updating according to an external clock is an over-simplification of biological reality. To alleviate this problem, instead of studying a MKN

$$x_i(t+1) = f_i(x_{i1}(t), x_{i2}(t), \dots, x_{iK_i}(t)) \quad , \quad i = 1, 2, \dots, N \quad , \quad (23)$$

with single-switch functions  $f_i$  derived from piecewise linear functions  $\overline{B}_i^{PW}$ , one could consider the corresponding system of piecewise linear differential equations

$$\frac{d}{dt}x_i(t) = \overline{B}_i^{PW}(x_{i1}(t), x_{i2}(t), \dots, x_{iK_i}(t)) - x_i(t) \quad , \quad i = 1, 2, \dots, N \quad . \quad (24)$$

Note that we can think of (23) as a simple Euler approximation of (24). The dynamical behavior of the continuous model has been shown to significantly differ from that of the Boolean network (Glass and Kauffman, 1973; Mestl et al., 1995; Edwards, 2000; de Jong et al., 2004; Mochizuki, 2005). In particular, the detection and analysis of chaos in systems of piecewise linear differential equations is still largely an open issue, although it was addressed in several, mostly numerical, studies (Mestl et al., 1997; Glass and Hill, 1998; Kappler et al., 2003).

Second, one could investigate the effect of different time-scales in the above system of piecewise linear differential equations (24) by multiplying the right-hand sides by time-scale parameters  $\tau_i$ . Several studies have shown that

the integration of different time-scales increases the robustness of biological systems (Gorban and Radulescu, 2007; Rojdestvenski et al., 1999). So far, however, authors almost exclusively treated the special case of equal time-scale parameters. The investigation of the general system will require the development of new analytical and numerical approaches. Still, we believe that the effort is worthwhile. Studying the effect of different time-scales on the self-organizing property of random networks will be interesting from both, a biological as well as a general systems-theoretic point of view.

Third, it remains to be investigated whether results about piecewise linear differential equations remain valid, if we do not approximate the sigmoidal functions in (17) by step functions. At least for sufficiently steep sigmoidals we expect this to be the case. However, from a biological point of view, the steepness is limited by the number of binding sites. This raises the question to what extent the "mathematical lower bound" and the "biological upper bound" can be reconciled.

### **Acknowledgements**

The authors wish to thank Florian Blöchl (Helmholtz Center, Munich) for critical reading of the manuscript and useful discussions. Funding by the Helmholtz Alliance on Systems Biology (project CoReNe) is gratefully acknowledged.

### **Appendix A.**

We prove that the expression for  $p_{S_i}$  from equation (1) has a unique global maximum if  $P_{S_i}$  is the discrete uniform distribution from (2). The existence

of a global maximum follows from the extreme value theorem. Now assume that  $p_1 \stackrel{\text{def}}{=} P_{S_i}(s_1) \neq p_2 \stackrel{\text{def}}{=} P_{S_i}(s_2)$ . Then, for  $p' \stackrel{\text{def}}{=} (p_1 + p_2)/2$  it holds that

$$\begin{aligned} 2p'(1-p') &= p_1 + p_2 - \frac{(p_1 + p_2)^2}{2} \\ &> p_1 + p_2 - \frac{(p_1 + p_2)^2}{2} - \frac{(p_1 - p_2)^2}{2} \\ &= p_1(1-p_1) + p_2(1-p_2). \end{aligned}$$

Hence, the distribution

$$P'_{S_i}(s) = \begin{cases} P_{S_i}(s) & s \notin \{s_1, s_2\} \\ p' & s \in \{s_1, s_2\} \end{cases}$$

yields a strictly larger value for  $p_{S_i}$ .

## Appendix B.

Observe that for  $o > 0$  the function  $O(o)$  is convex and monotonously increasing. To prove this, it suffices to show that  $o^K + (1 - o^K)(1 - p_S)$  is convex and monotonously increasing. This can easily be verified by computing the first and second derivatives,  $K o^{K-1} - (1 - p_S) K o^{K-1} = K p_S o^{K-1} \geq 0$  and  $K(K-1) p_S o^{K-2} \geq 0$ , respectively. Moreover, note that  $p_S < 1$  implies  $O(0) > 0$ . Hence, there are two possibilities: Either the fixed point  $o^*$  is (globally) attractive and no further fixed points exist, or  $o^*$  is repellent and  $O$  has exactly one additional (attractive) fixed point. These two possibilities are visualized in Figure B.3. We observe that for the blue curve the fixed point  $o^*$  is attractive (even globally), while for the green curve it is repellent. In the latter case we have an additional fixed point, which attracts  $[0, 1)$ .

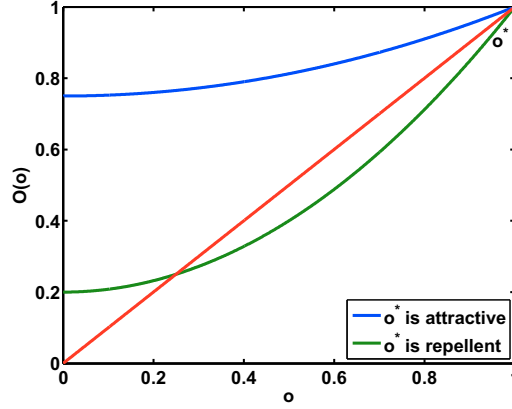


Figure B.3: Schematic plots of  $O(o)$  from (4) for a frozen (blue) and a chaotic (green) KN.

The linear stability analysis of (4) about  $o^*$  yields:

$$\begin{aligned}
 \left. \frac{dO}{do} \right|_{o^*} &= \sum_{K=1}^{K_{\max}} P_{\text{in}}(K) \sum_{S=2}^{S_{\max}} P_{\text{nos}}(S) [K - K(1 - p_S)] \\
 &= \sum_{K=1}^{K_{\max}} P_{\text{in}}(K) \sum_{S=2}^{S_{\max}} P_{\text{nos}}(S) K p_S \\
 &= \sum_{K=1}^{K_{\max}} P_{\text{in}}(K) K \bar{p} \\
 &= \bar{p} \bar{K} \begin{cases} < 1 & \text{ordered regime} \\ = 1 & \text{critical boundary} \\ > 1 & \text{chaotic regime .} \end{cases}
 \end{aligned}$$

### Appendix C.

We substitute (14) in (5):

$$\bar{K} [2w(1-w) + (1-w)^2(1-\mu_{-1})] \begin{cases} < 1 & \text{ordered regime} \\ = 1 & \text{critical boundary} \\ > 1 & \text{chaotic regime .} \end{cases}$$

This is equivalent to

$$-\bar{K}(1+\mu_{-1})w^2 + 2\bar{K}\mu_{-1}w + \bar{K}(1-\mu_{-1}) - 1 \begin{cases} < 0 & \text{ordered regime} \\ = 0 & \text{critical boundary} \\ > 0 & \text{chaotic regime .} \end{cases} \quad (\text{C.1})$$

The discriminant of this quadratic equation is given by

$$\mathcal{D} = 4\bar{K}(\bar{K} - (1 + \mu_{-1}))$$

and

$$\mathcal{D} \begin{cases} < 0 & \text{if } \bar{K} < 1 + \mu_{-1} \\ = 0 & \text{if } \bar{K} = 1 + \mu_{-1} \\ > 0 & \text{if } \bar{K} > 1 + \mu_{-1} . \end{cases}$$

From (C.1) it follows that for  $\bar{K} < 1 + \mu_{-1}$  the system is always frozen. To see this, note that  $-\bar{K}(1 + \mu_{-1}) < 0$ . For  $\bar{K} \geq 1 + \mu_{-1}$  the quadratic equation from (C.1) has two (possibly coinciding) solutions

$$w_{\frac{1}{2}}^{\text{crit}}(\bar{K}, \mu_{-1}) = \frac{1}{\mu_{-1} + 1} \left( \mu_{-1} \mp \sqrt{1 - \frac{\mu_{-1} + 1}{\bar{K}}} \right)$$

First, observe that for  $\mu_{-1} > 0$ ,  $w_{\frac{1}{2}}^{\text{crit}} < 1$  is always a valid solution. Moreover, we have

$$w_{\frac{1}{2}}^{\text{crit}} > 0 \Leftrightarrow \sqrt{1 - \frac{\mu_{-1} + 1}{\bar{K}}} < \mu_{-1} \Leftrightarrow \bar{K} < \frac{1}{1 - \mu_{-1}},$$



where in the case  $\mu_{-1} = 1$ , " $\bar{K} < \infty$ " is true for any  $\bar{K}$ . Observe that for  $\mu_{-1} < 1$ ,

$$\frac{1}{1 - \mu_{-1}} > 1 + \mu_{-1},$$

so there always is a range for  $\bar{K}$ , in which (C.1) has two distinct solutions in  $(0, 1)$ . Once more considering that in (C.1)  $-\bar{K}(1 + \mu_{-1}) < 0$ , we finally obtain

$$-\bar{K}(1 + \mu_{-1})w^2 + 2\bar{K}\mu_{-1}w + \bar{K}(1 - \mu_{-1}) - 1 \begin{cases} < 0, & w < w_1^{\text{crit}} \text{ or } w > w_2^{\text{crit}} \\ = 0, & w = w_1^{\text{crit}} \text{ or } w = w_2^{\text{crit}} \\ > 0, & w_1^{\text{crit}} < w < w_2^{\text{crit}}. \end{cases}$$

#### Appendix D.

Using  $\mu_{-1} = 1/(\bar{S} - 1)$  we can rewrite the classification from section 3.1 in terms of  $\bar{S}$ . We also write  $w^{\text{crit}}(\bar{K}, \bar{S})$  instead of  $w^{\text{crit}}(\bar{K}, \mu_{-1})$ .

(a) For  $\bar{K} < \bar{S}/(\bar{S} - 1)$  the system is always frozen.

(b) For  $\bar{K} = \bar{S}/(\bar{S} - 1)$  the system is critical if and only if

$$w = w^{\text{crit}}\left(\frac{\bar{S}}{\bar{S} - 1}, \bar{S}\right) = \frac{1}{\bar{S}},$$

i.e.  $P_{\bar{S}}$  is the uniform distribution, otherwise it is frozen.

(c) For  $\bar{S} = 2$  and  $\bar{K} > \bar{S}/(\bar{S} - 1)$  the critical boundary is described by

$$w_{\frac{1}{2}}^{\text{crit}}(\bar{K}, 2) = \frac{1}{2} \left[ 1 \mp \sqrt{1 - \frac{2}{\bar{K}}} \right] \in (0, 1). \quad (\text{D.1})$$

For

$$w_1^{\text{crit}}(\bar{K}, 2) < w < w_2^{\text{crit}}(\bar{K}, 2)$$

the system is chaotic, otherwise it is frozen.

- (d) For  $\bar{S} > 2$  and  $(\bar{S} - 1)/(\bar{S} - 2) > \bar{K} > \bar{S}/(\bar{S} - 1)$  the critical boundary is described by

$$w_{\frac{1}{2}}^{\text{crit}}(\bar{K}, \bar{S}) = \frac{1}{\bar{S}} \left( 1 \mp \sqrt{1 - \frac{\bar{S}}{\bar{K}} [(2 - \bar{S}) \bar{K} + \bar{S} - 1]} \right) \in (0, 1). \quad (\text{D.2})$$

For

$$w_1^{\text{crit}}(\bar{K}, \bar{S}) < w < w_2^{\text{crit}}(\bar{K}, \bar{S}) \in (0, 1)$$

the system is chaotic, otherwise it is frozen. For  $\bar{K} = (\bar{S} - 1)/(\bar{S} - 2)$ ,  $w_1^{\text{crit}}(\bar{K}, \bar{S}) = 0$  is not a valid solution for  $w$  as we require  $w \in (0, 1)$ . Note that for  $w = 0$ ,  $P_{\bar{S}}$  from (13) would essentially become a uniform distribution on  $1, 2, \dots, \bar{S} - 1$ . We could replace  $\bar{S}$  by  $\bar{S}' = \bar{S} - 1$  and choose a uniform distribution for  $P_{\bar{S}'}$ . This new system falls into category (b) and we find that it is in fact critical.

- (e) For  $\bar{S} > 2$  and  $\bar{K} \geq (\bar{S} - 1)/(\bar{S} - 2)$  the critical boundary is described by

$$w^{\text{crit}}(\bar{K}, \bar{S}) = \frac{1}{\bar{S}} \left[ 1 + \sqrt{1 - \frac{\bar{S}}{\bar{K}} [(2 - \bar{S}) \bar{K} + \bar{S} - 1]} \right] \in (0, 1), \quad (\text{D.3})$$

cf. Solé et al. (2000), for larger values of  $w$  the system is frozen, otherwise it is chaotic.

## Appendix E.

Let us assume a situation as outlined at the beginning of section 4.2. We begin our computations without any assumptions regarding cooperative effects. Hence, there are  $2^{\sum_{k=1}^K n_k}$  different accessible promotor configurations

(*microstates*); we describe them by Boolean vectors  $\omega$  of length  $\sum_{k=1}^K n_k$ , the  $b$ -th component  $\omega(b)$  indicating if the respective binding site is occupied. (The reader is kindly asked to ignore that this definition of  $\omega$  conflicts with the one given in the main text, we will resolve this discrepancy below.) For each microstate  $\omega$  we denote the number of binding sites occupied by regulator  $x_k$  by  $\#\omega(k)$ .

Each configuration  $\omega$  has a certain *Gibbs free energy*  $G_\omega$  and the probability of a certain microstate  $\omega_0$  is given by

$$\pi_{\omega_0} = \frac{\exp(-\Delta G_{\omega_0}/RT) \prod_{k=1}^K [x_k]^{\#\omega_0(k)}}{\sum_{\omega} \exp(-\Delta G_{\omega}/RT) \prod_{k=1}^K [x_k]^{\#\omega(k)}}, \quad (\text{E.1})$$

where  $\Delta$  signifies energy differences with respect to the state  $(0, 0, \dots, 0)$ ,  $R$  is the gas constant,  $T$  the temperature, and brackets indicate concentrations of free ligand. Since we may assume the number of ligands to be larger than the number of promoters by several orders of magnitude, we can neglect the difference between concentration of free ligand and total ligand concentration.

Formula (E.1) can also be written in terms of the *association constants*  $A_\omega$ , which are related to the free energies via  $\Delta G_\omega = -RT \ln A_\omega$ . Observe that  $\Delta G_{(0,0,\dots,0)} = 0$  implies  $A_{(0,0,\dots,0)} = 1$ . We obtain

$$\pi_{\omega_0} = \frac{A_{\omega_0} \prod_{k=1}^K [x_k]^{\#\omega_0(k)}}{\sum_{\omega} A_{\omega} \prod_{k=1}^K [x_k]^{\#\omega(k)}}. \quad (\text{E.2})$$

Let us begin by considering the case  $K = 1$ . We denote the free energy of binding a protein to the  $b$ -th binding site by  $\Delta G_{e_b}$ , where  $e_b$  is the  $b$ -th unit vector of length  $n_1$ . The free energy  $\Delta G_\omega$  of a microstate  $\omega$  can then be written as

$$\Delta G_\omega = \sum_{b|\omega(b)=1} \Delta G_{e_b} + \Delta G_\omega^{\text{hom}}, \quad (\text{E.3})$$

i.e. as the sum of the free energies  $\Delta G_{e_b}$  of the single protein-DNA bindings and the free energy  $\Delta G_{\omega}^{\text{hom}}$  of interactions between the proteins (*homotypic cooperativity*). Clearly,  $\Delta G_{e_b}^{\text{hom}} = 0$  for all unit vectors  $e_b$ .

For simplification, one commonly assumes that either no protein or  $n_1$  proteins are bound to the promotor, i.e. one discards all microstates except  $\vec{0} = (0, 0, \dots, 0)$  and  $\vec{1} = (1, 1, \dots, 1)$ . A possible justification of this assumption is a strong positive homotypic cooperativity. Indeed, according to (E.1) and (E.3) we can neglect all  $\pi_{\omega}$  except  $\pi_{\vec{0}}$  and  $\pi_{\vec{1}}$  if  $-\Delta G_{\vec{1}}^{\text{hom}}$  is sufficiently large. This issue is further discussed in section 6.4.

Equation (E.2) then becomes

$$\pi_{\vec{0}} = \frac{1}{1 + A_{\vec{1}} [x_1]^{n_1}}, \quad \pi_{\vec{1}} = \frac{A_{\vec{1}} [x_1]^{n_1}}{1 + A_{\vec{1}} [x_1]^{n_1}},$$

or, in terms of the *dissociation constant*  $D_{\vec{1}} = 1/A_{\vec{1}}$ ,

$$\pi_{\vec{0}} = \frac{D_{\vec{1}}}{D_{\vec{1}} + [x_1]^{n_1}}, \quad \pi_{\vec{1}} = \frac{[x_1]^{n_1}}{D_{\vec{1}} + [x_1]^{n_1}}.$$

These are the well-known negative and positive Hill functions (Hill, 1910), which are often used to model transcriptional gene regulation. Typically, they are stated in terms of the *microscopic dissociation constant*  $D_{\vec{1}}^{\text{micro}} = \sqrt[n_1]{D_{\vec{1}}}$ , which indicates the concentration level for half-maximal activation. The basic idea behind discrete models of gene regulatory networks is to approximate the sigmoidal Hill functions by Heaviside step functions, cf. Figure 2C.

Let us now consider the case of general  $K$ . For simplification, we again assume that each transcription factor  $x_k$ ,  $1 \leq k \leq K$ , can bind only as a polymer consisting of  $n_k$  monomers. This limits the number of accessible microstates to  $2^K$ ; by abuse of notation we denote them again by vectors  $\omega \in \{0, 1\}^K$ , the  $k$ -th component  $\omega(k)$  indicating whether or not the  $k$ -th

regulator polymer is bound. With this new definition of  $\omega$ , which now agrees with the one given in the main text, equation (E.2) reads

$$\pi_{\omega_0} = \frac{A_{\omega_0} \prod_{k|\omega_0(k)=1} [x_k]^{n_k}}{\sum_{\omega} A_{\omega} \prod_{k|\omega(k)=1} [x_k]^{n_k}}. \quad (\text{E.4})$$

In general, it holds that

$$\Delta G_{\omega} = \sum_{k|\omega(k)=1} \Delta G_{e_k} + \Delta G_{\omega}^{\text{het}},$$

where  $\Delta G_{e_k}$  is the free energy of binding the  $k$ -th regulator polymer and  $\Delta G_{\omega}^{\text{het}}$  is the free energy of interactions between the different regulators (*heterotypic cooperativity*). In the following, we neglect heterotypic cooperativity, all  $\Delta G_{\omega}^{\text{het}} = 0$ , for a discussion see section 6.4. In terms of the association constants this implies

$$A_{\omega} = \prod_{k|\omega(k)=1} A_{e_k}. \quad (\text{E.5})$$

Hence, the denominator in (E.4) factorizes

$$\sum_{\omega} A_{\omega} \prod_{k|\omega(k)=1} [x_k]^{n_k} = \sum_{\omega} \prod_{k|\omega(k)=1} A_{e_k} [x_k]^{n_k} = \prod_{k=1}^K (1 + A_{e_k} [x_k]^{n_k}). \quad (\text{E.6})$$

Using (E.5) and (E.6) we can rewrite (E.4) as

$$\begin{aligned} \pi_{\omega_0} &= \frac{\prod_{k|\omega_0(k)=1} A_{e_k} [x_k]^{n_k}}{\prod_{k=1}^K (1 + A_{e_k} [x_k]^{n_k})} \\ &= \prod_{k|\omega_0(k)=1} \frac{A_{e_k} [x_k]^{n_k}}{1 + A_{e_k} [x_k]^{n_k}} \cdot \prod_{k|\omega_0(k)=0} \frac{1}{1 + A_{e_k} [x_k]^{n_k}} \\ &= \prod_{k|\omega_0(k)=1} h_+([x_k], n_k, A_{e_k}) \cdot \prod_{k|\omega_0(k)=0} h_-([x_k], n_k, A_{e_k}). \end{aligned}$$

Replacing  $A_{e_k}$  by  $A_k$  and dropping the subscript 0 we obtain (15).

## Appendix F.

Let us consider the  $i$ -th node,  $1 \leq i \leq N$ , and its  $k$ -th input,  $1 \leq k \leq K_i$ . The probability  $q_{S_{ik}}$  that for two different values  $x_{ik} \neq x'_{ik}$  of this input we have  $h_{\text{step}}(x_{ik}, \theta) \neq h_{\text{step}}(x'_{ik}, \theta)$ , where  $\theta \in \Sigma_{ik} \setminus \{0\}$  is a randomly chosen threshold, is given by

$$q_{S_{ik}} = \frac{1}{S_{ik} - 1} \sum_{\theta=1}^{S_{ik}-1} 2 \sum_{s=0}^{\theta-1} \left[ P_{S_{ik}}(s) \frac{\sum_{s'=\theta}^{S_{ik}-1} P_{S_{ik}}(s')}{1 - P_{S_{ik}}(s)} \right]. \quad (\text{F.1})$$

To see this, think of  $x_{ik}$  and  $x'_{ik}$  as i.i.d. random variables distributed according to  $P_{S_{ik}}$ . The bracketed expression is the product of the probability for " $x_{ik} = s$ " and the conditional probability for " $x'_{ik} \geq \theta$ " given " $x'_{ik} \neq s$ ". The sum over  $s$  thus gives the probability for " $x_{ik} < \theta$  and  $x'_{ik} \geq \theta$ " given " $x_{ik} \neq x'_{ik}$ ". Twice this sum is the probability that  $x_{ik}$  and  $x'_{ik}$  lie on different sides of  $\theta$  and thus give different values when plugged into  $h_{\text{step}}(\bullet, \theta)$ . Computing the weighted sum over all non-degenerate thresholds  $\theta$  finally yields the above formula (F.1).

In the situation of section 5.2 formula (F.1) simplifies to

$$q_{\bar{S}} = \bar{q} = \frac{1}{\bar{S} - 1} \sum_{\theta=1}^{\bar{S}-1} 2 \frac{\theta \bar{S} - \theta}{\bar{S} \bar{S} - 1} = \frac{1 \bar{S} + 1}{3 \bar{S} - 1}. \quad (\text{F.2})$$

It is interesting to compare this to the continuous situation, where the two input values and the threshold are drawn from some continuous interval, w.l.o.g. the unit interval  $[0, 1]$ . Intuitively one might suspect that the probability  $q^{\text{cont}}$  for the two input values to lie on different sides of the threshold is  $1/2$ . However,  $q^{\text{cont}}$  computes to

$$q^{\text{cont}} = \int_0^1 2x(1-x) dx = \frac{1}{3}$$

and agrees with the limit of  $\bar{q}$  from (F.2) as  $\bar{S} \rightarrow \infty$ .

## References

- Albert, R., Othmer, H. G., 2003. The topology of the regulatory interactions predicts the expression pattern of the segment polarity genes in *Drosophila melanogaster*. *J Theor Biol* 223, 1–18.
- Aldana, M., 2003. Boolean dynamics of networks with scale-free topology. *Physica D* 185 (1), 45–66.
- Aldana, M., Coppersmith, S., Kadanoff, L. P., 2003. Boolean dynamics with random couplings. *Perspectives and Problems in Nonlinear Science*, 23–89.
- Ballesteros, F., Luque, B., 2005. Order-disorder phase transition in random-walk networks. *Phys Rev E* 71 (3), 31104.
- Balleza, E., Alvarez-Buylla, E., Chaos, A., Kauffman, S., Shmulevich, I., Aldana, M., 2008. Critical dynamics in genetic regulatory networks: examples from four kingdoms. *PLoS One* 3 (6).
- Bastolla, U., Parisi, G., 1996. Closing probabilities in the Kauffman model: an annealed computation. *Physica D* 98 (1), 1–25.
- Buchler, N., Gerland, U., Hwa, T., 2003. On schemes of combinatorial transcription logic. *Proc Natl Acad Sci U S A* 100 (9), 5136–5141.
- de Jong, H., Gouzé, J., Hernandez, C., Page, M., Sari, T., Geiselmann, J., 2004. Qualitative simulation of genetic regulatory networks using piecewise-linear models. *Bull Math Biol* 66 (2), 301–340.
- Derrida, B., Pomeau, Y., 1986. Random networks of automata: a simple annealed approximation. *Europhys. Lett.* 1 (2), 45–49.

- Derrida, B., Stauffer, D., 1986. Phase transitions in two dimensional Kauffman cellular automata. *Europhys. Lett.* 2 (10), 739–745.
- Edwards, R., 2000. Analysis of continuous-time switching networks. *Physica D* 146 (1–4), 165–199.
- Faisal, S., Lichtenberg, G., Werner, H., 2008. Polynomial models of gene dynamics. *Neurocomputing* 71 (13–15), 2711–2719.
- Fauré, A., Naldi, A., Lopez, F., Chaouiya, C., Ciliberto, A., Thieffry, D., 2009. Modular logical modelling of the budding yeast cell cycle. *Mol BioSyst* 5, 1787–1796.
- Flyvbjerg, H., 1988. An order parameter for networks of automata. *J. Phys. A: Math. Gen.* 21, L955–L960.
- Glass, L., Hill, C., 1998. Ordered and disordered dynamics in random networks. *Europhys. Lett.* 41, 599–604.
- Glass, L., Kauffman, S. A., 1973. The logical analysis of continuous, non-linear biochemical control networks. *J Theor Biol* 39 (1), 103–129.
- Gorban, A., Radulescu, O., 2007. Dynamical robustness of biological networks with hierarchical distribution of time scales. *IET Syst. Biol.* 1 (4), 238–246.
- Harris, S., Sawhill, B., Wuensche, A., Kauffman, S. A., 2002. A model of transcriptional regulatory networks based on biases in the observed regulation rules. *Complexity* 7 (4), 23–40.



- Hilhorst, H., Nijmeijer, M., 1987. On the approach of the stationary state in Kauffman's random Boolean network. *J. Phys.* 48 (2), 185–191.
- Hill, A., 1910. The possible effects of the aggregation of the molecules of haemoglobin on its dissociation curves. *J Physiol* 40, 4–7.
- Kappler, K., Edwards, R., Glass, L., 2003. Dynamics in high-dimensional model gene networks. *Signal Processing* 83 (4), 789–798.
- Kauffman, S. A., 1969. Metabolic stability and epigenesis in randomly constructed genetic nets. *J Theor Biol* 22 (3), 437–467.
- Kauffman, S. A., 1993. *The origins of order: Self organization and selection in evolution.* Oxford University Press, USA.
- Kauffman, S. A., Peterson, C., Samuelsson, B., Troein, C., 2003. Random Boolean network models and the yeast transcriptional network. *Proc Natl Acad Sci U S A* 100 (25), 14796–14799.
- Kauffman, S. A., Peterson, C., Samuelsson, B., Troein, C., 2004. Genetic networks with canalizing Boolean rules are always stable. *Proc Natl Acad Sci U S A* 101 (49), 17102–17107.
- Kendall, M., Stuart, A., Ord, J., 1987. *Kendall's advanced theory of statistics.* Oxford University Press, Inc. New York, NY, USA.
- Li, F., Long, T., Lu, Y., Quyang, Q., Tang, C., 2004. The yeast cell-cycle network is robustly designed. *Proc Natl Acad Sci U S A* 101 (14), 4781–4786.

- Luque, B., Ballesteros, F., 2004. Random walk networks. *Physica A* 342 (1-2), 207–213.
- Mestl, T., Bagley, R., Glass, L., 1997. Common chaos in arbitrarily complex feedback networks. *Phys. Rev. Lett.* 79 (4), 653–656.
- Mestl, T., Plahte, E., Omholt, S. W., 1995. A mathematical framework for describing and analysing gene regulatory networks. *J Theor Biol* 176 (2), 291–300.
- Mochizuki, A., 2005. An analytical study of the number of steady states in gene regulatory networks. *J Theor Biol* 236 (3), 291–310.
- Raeymaekers, L., 2002. Dynamics of Boolean Networks Controlled by Biologically Meaningful Functions. *J Theor Biol* 218 (3), 331–341.
- Rohlf, T., Bornholdt, S., 2002. Criticality in random threshold networks: annealed approximation and beyond. *Physica A* 310 (1–2), 245–259.
- Rojdestvenski, I., Cottam, M., Park, Y., Öquist, G., 1999. Robustness and time-scale hierarchy in biological systems. *BioSystems* 50 (1), 71–82.
- Saez-Rodriguez, J., Simeoni, L., Lindquist, J. A., Hemenway, R., Bommhardt, U., Arndt, B., Haus, U.-U., Weismantel, R., Gilles, E. D., Klamt, S., Schraven, B., 2007. A logical model provides insights into t cell receptor signaling. *PLoS Comput Biol* 3 (8), e163.
- Sánchez, L., Thieffry, D., 2001. A logical analysis of the *Drosophila* gap-gene system. *J Theor Biol* 211 (2), 115–141.

- Segal, E., Raveh-Sadka, T., Schroeder, M., Unnerstall, U., Gaul, U., 2008. Predicting expression patterns from regulatory sequence in *Drosophila* segmentation. *Nature* 451 (7178), 535–540.
- Setty, Y., Mayo, A. E., Surette, M. G., Alon, U., 2003. Detailed map of a cis-regulatory input function. *Proc Natl Acad Sci U S A* 100 (13), 7702–7707.
- Sneppen, K., Zocchi, G., 2005. *Physics in molecular biology*. Cambridge University Press.
- Solé, R. V., Luque, B., Kauffman, S. A., 2000. Phase transitions in random networks with multiple states. Working Papers 00-02-011, Santa Fe Institute.  
URL <http://ideas.repec.org/p/wop/safiwp/00-02-011.html>
- Szejka, A., Mihaljev, T., Drossel, B., 2008. The phase diagram of random threshold networks. *New J. Phys.* 10, 063009.
- Thomas, R., 1991. Regulatory networks seen as asynchronous automata: A logical description. *J Theor Biol* 153 (1), 1–23.
- Weiss, J., 1997. The Hill equation revisited: uses and misuses. *The FASEB Journal* 11 (11), 835–841.
- Wittmann, D. M., Blöchl, F., Trümbach, D., Wurst, W., Prakash, N., Theis, F. J., 11 2009a. Spatial analysis of expression patterns predicts genetic interactions at the mid-hindbrain boundary. *PLoS Comput Biol* 5 (11), e1000569.

Wittmann, D. M., Krumsiek, J., Saez-Rodriguez, J., Lauffenburger, D. A., Klamt, S., Theis, F. J., 2009b. Transforming Boolean Models to Continuous Models: Methodology and Application to T-Cell Receptor Signaling. *BMC Syst Biol* 3, 98.

Accepted manuscript