The Pennsylvania State University

The Graduate School

Eberly College of Science

ANALYSIS AND LOGICAL MODELING OF BIOLOGICAL SIGNALING TRANSDUCTION NETWORKS

A Dissertation in Physics

by

Zhongyao Sun

Submitted in Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

May 2015

The dissertation of Zhongyao Sun was reviewed and approved* by the following:

Réka Albert

Distinguished Professor of Physics and Biology

Dissertation Advisor

Chair of Committee

Dezhe Z. Jin

Associate Professor of Physics

Jorge O. Sofo

Professor of Physics

Professor of Materials Science and Engineering

John Fricks

Associate Professor of Statistics

Nitin Samarth

Professor of Physics

Department Head

*Signatures are on file in the Graduate School.

ABSTRACT

The study of network theory and its application span across a multitude of seemingly disparate fields of science and technology: computer science, biology, social science, linguistics, etc. It is the intrinsic similarities embedded in the entities and the way they interact with one another in these systems that link them together.

In this dissertation, I present from both the aspect of theoretical analysis and the aspect of application three projects, which primarily focus on signal transduction networks in biology. In these projects, I assembled a network model through extensively perusing literature, performed model-based simulations and validation, analyzed network topology, and proposed a novel network measure. The application of network modeling to the system of stomatal opening in plants revealed a fundamental question about the process that has been left unanswered in decades. The novel measure of the redundancy of signal transduction networks with Boolean dynamics by calculating its maximum node-independent elementary signaling mode set accurately predicts the effect of single node knockout in such signaling processes. The three projects as an organic whole advance the understanding of a real system as well as the behavior of such network models, giving me an opportunity to take a glimpse at the dazzling facets of the immense world of network science.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 1

Introduction

Every entity in our universe is connected, related, or correlated with one another. From the constant ebbs and flows of energy and matter, the perpetual interaction between species and people, to the intricate way concepts and words are woven together forming our languages and the knowledge itself, such relationships form and are therefore best represented by networks. Network modeling is one of the most natural ways to approach systems the components of which interact with each other in a complex manner. It permeates numerous spheres of science and technology: cellular biology, ecology, social sciences, logistics, etc., making it one of the most popular and fruitful methodologies for researchers. In addition to the network structure, one often desires to describe the dynamics of mass or information transfer through the network. Dynamic models characterize the nodes by states (e.g. concentration or activity), and the states of the nodes change in time according to the interactions encapsulated in the network. My doctoral study focused on network-based discrete dynamic modeling of biological signaling transduction systems and theoretical analysis of such models.

The reason I chose biological signaling processes as the subject of my study is manifold: first and foremost, such processes can be and are best represented as networks, due to the enormous amount of players involved in such processes and the complexity of their interrelations. Modeling such systems with networks is an intriguing and fascinating topic. Secondly, signaling processes underlie almost all biological behaviors on a microscopic scale, e.g. cell response to external stimuli such as pathogen or light. Signaling processes are vitally important to not only the survival of the single cell, but also the wholeness of individuals. Thirdly, what have been observed for such processes is that it is the relation of a concentration relative to a threshold value that matters rather than the absolute quantity itself. Such property of signaling processes provides the opportunity to apply Boolean dynamics, which facilitates the construction and analysis of such network models.

Chapter 2 offers a brief introduction to the process from the assembly of networks based on literature to the implementation of a specific form of dynamics—Boolean dynamics—based on the network structure. The necessary ingredients of Boolean dynamic modeling are introduced, and three examples of its successful application are given.

Chapter 3, 4 and 5 present several facets of the work I was engaged in. In chapter 3, a network model was constructed with discrete dynamics for the biological process known as light-induced stomatal opening and its regulation by abscisic acid. The methodology of discrete dynamics is a natural extension of Boolean dynamics, in which both the values a node can assume and the means these values are regulated over time takes on a greater degree of freedom. The application of dynamic modeling offers a multitude of novel insights into and predictions about the behavior of the system. For instance, it revealed a basic question about the relationship between red light-induced stomatal opening and abscisic acid that had been overlooked for decades.

Chapter 4 and 5 detail my original work on theoretical analysis conducted under the Boolean network framework and the concept of elementary signaling mode. Chapter 4 develops a measure of network redundancy and robustness based on node-independent elementary signaling modes. The novel measure is subsequently tested on a number of existing Boolean network models and is demonstrated to precisely predict the nodes of critical importance for a signaling cascade. Chapter 5 introduces a procedure of solving for the attractor of Boolean networks without carrying out dynamic simulations. The attractors of Boolean models correspond to the time-invariant behavior of real systems, e.g. cell apoptosis, cell type, phenotype, etc. It is therefore of high importance that these attractors can be identified without resorting to dynamic simulations, which, as the number of nodes in the system increases, becomes increasingly tedious and time-consuming, to the point where it is no longer tenable. The method formulated defines a succinct as well as precise process with which one can identify all attractors of a Boolean network, offering future researchers using Boolean network models a powerful tool.

Chapter 2

Boolean Dynamic Modeling of Biological Systems

2.1 Overview

Network structures permeate every sphere of cellular biology. The vast number of intra- and extracellular processes and interactions that form a complex web of mass-, energy- and signal transfer can intrinsically be described in network language. The nodes of a biological network are cellular components such as genes, RNAs, proteins and small molecules and the edges are reactions, interactions, regulatory or synthetic relationships among components. The process of transcribing coding DNA into mRNAs is either promoted or suppressed by transcription factors; the totality of these transcriptional processes can be integrated into gene regulatory networks (1). Similarly, diverse interactions amongst proteins, or the biochemical reactions in cellular metabolism, can be readily depicted in network language (2-4). In a network representation, the elements of the system are represented by network nodes, and the interactions among elements as edges. The wealth of data and the affinity between network science and biology make network modeling of biological systems not only viable, but also powerful and uniquely useful.

In addition to the network structure (that is, the nodes and edges), it is often desired to be able to describe the dynamics of mass or information transfer through the network. Dynamic models characterize the nodes by states (e.g. concentration or activity), and the states of the nodes change in time according to the interactions encapsulated in the network. Continuous dynamic models use sets of differential equations to capture the detailed variation of concentrations of key substances (5) in the system. Despite the recent phenomenal growth of computing power, such approaches become practically impossible to implement when the number of nodes in a system

reaches more than one hundred. Difficulties are also posed for parameter estimation in large-scale systems when there is insufficient temporal data. The aim to circumvent these difficulties makes discrete dynamic modeling such as Boolean networks (6-13) and Petri nets (14-16) particularly useful. Vast amounts of qualitative knowledge of regulatory relationships between cellular components have been either experimentally measured or inferred from biological evidence. The compilation of such knowledge forms the basis for network reconstruction, which then allows discrete modeling of a system, bypassing the obstacles posed by parameter estimation. The model can be tested against experimental evidence and refined iteratively. Discrete dynamic models generate insightful analysis about the interweaving of system components, the cascade of information flow, and enable *in silico* node knockout experiments that produce informative predictions about the system. Discrete dynamic models have been successfully implemented in numerous biological systems, facilitating the study and deepening understanding of biological processes such as flower development (17, 18), the yeast cell cycle (8, 9, 19), *Drosophila* embryonic development (20-24), hormone signaling in plants (10, 25), the immune response (11), and T cell signaling and differentiation (12, 13, 26-28).

Boolean network modeling of biological systems is the simplest of the discrete dynamic models. Each node can have one of two discrete states, namely 0 and 1, instead of a continuously varying concentration. 0 or *OFF* means that the element represented by the node is inactive (e.g. an enzyme or transcription factor), or has a below-threshold concentration (e.g. a small molecule). 1 or *ON* represents the opposite, which is active or an above-threshold concentration (6, 7). The state of a system that has N nodes is therefore represented by an N-dimensional vector with each value being 0 or 1. The state space of such system contains a total of 2^N states. In order to carry out dynamic simulations, time is usually discretized into time steps. As time evolves, starting from

one initial state vector, either predetermined or randomly generated, the system state vector can traverse the state space, reaching different parts of the space. It is not necessary that a dynamic trajectory traverse all 2^N states in the space; indeed the observed trajectories converge into a stationary state or set of states after a much smaller number of state transitions.

In this chapter, we first explain the correspondence between Boolean networks and biological systems, then introduce the basic concepts of Boolean network modeling and elaborate in a step-by-step fashion the modeling procedure, starting from network reconstruction based on biological information compilation to model validation. We will present two successful applications of this methodology in cellular biology (11, 12).

2.2 Boolean Networks and Biological Systems

Here we introduce the basics of Boolean algebra and its connection with biological systems. In order to summarize all available information and graphically represent the system we are studying, the elements of the system, such as genes or proteins, become the nodes of the network. Nodes that represent a phenomenon or a certain biological result can also be included, for instance Stomatal Closure, or Apoptosis.

Edges are drawn between biologically or chemically related nodes to represent the relationship between them. Since the flow of signals and reaction fluxes are directional, the edges in the network will each have a direction that is consistent with biology; edges may also be characterized by a sign (+/—) that denotes the property of that edge: positive sign for activation and negative for inhibition. For example a common representation of the synthesis reaction C + D →E connects the two reactants with the product of the reaction as in Figure 2-1A.

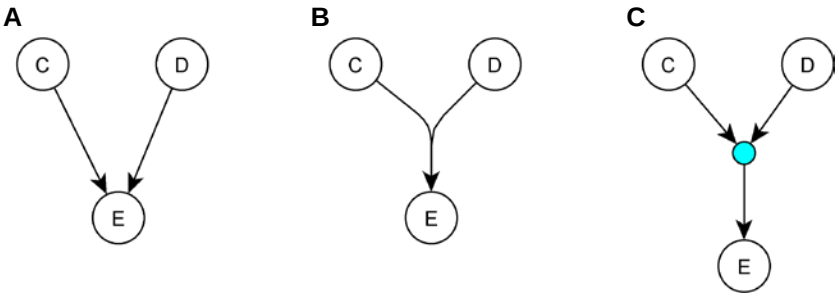


Figure 2-1. Three different ways to represent the same synthesis reaction relationship between 3 nodes. (A) Two separate edges directed from the reactants to the product. (B) Two edges from the reactants first merged together and then directed at the product. (C) An intermediate node denoting the synergy between the two reactants is added. The reactants are connected to this intermediary node, which is in turn connected to the product.

This representation does not, however, directly reflect the fact that C and D are both required for the reaction to take place. Thus the network must be complemented with rules that specify the ways in which all upstream components are combined. A natural and economical method is to use the Boolean operators *AND, OR, NOT*. A combination of these operators can describe most possible relationships or reactions between substances and components in a biological system. The logic dependence underlying a synthesis reaction can be described by the Boolean operator *AND*, so the reaction $C + D \rightarrow E$ becomes $C\ AND\ D = E^*$ in Boolean language, where for simplicity the node names stand for the state of the node. Similarly, if a certain result can stem from multiple causes, e.g. photosynthesis can be carried out under either blue or red light, then the Boolean equation denoting that will be Blue_light *OR* Red_light = Photosynthesis*. When a certain element or activity is negatively regulated by another, the *NOT* rule is used. The asterisks on the two example Boolean equations indicate that the specific processes to generate the product on the right hand side take a certain amount of time to complete. We return to this point in section 2.3.3.

The additional information contained in the Boolean rules (e.g. the conditionality or independence of two edges incident on the same node) can be integrated into the network for a more complete representation. Merging edges may be used to represent the Boolean *AND* relation (see Figure 1(b)), and separate edges for *OR* relations (see Figure 1(a))(29). There is also a third choice, which involves an intermediate node, if the Boolean rule is *AND* (see Figure 1(c)). Such a node would not exist if the rule were Boolean *OR* (21, 30, 31). So far there is no universal 'standard' for how to map a system of complex relations onto a graph. The most appropriate choice is the one that best facilitates the analysis of the particular system under study.

2.3 Boolean Network Modeling

As mentioned in the introduction, a dynamic model characterizes a system's behavior over time. In Boolean dynamic modeling, specifically, both the system's status, and time are discretized. The state of a node can be 0 or 1, making the state of a system of N nodes an N-dimensional vector of 0's and 1's. A continuous time stream over a certain period (e.g. an experiment) is represented by a series of steps, which are an abstract representation of important time points at which biochemical events are taking place. The state of the system at a time step is determined by its predecessor state (or sometimes several predecessor states at earlier time steps) through what are called Boolean transfer functions. The calculation of a node state at a time step based on system state(s) at earlier step(s) is called 'updating'. Depending on the updating scheme that is used, a number of nodes, ranging from 1 to N, are updated, thus obtaining the system state at a new (future) time step. Given below is an example of a general expression of a Boolean transfer function of a certain node i. Suppose the state of node i at time step t is denoted as $V_{i,t}$.The transfer function through which $V_{i,t}$is calculated is given as follows:

$$V_{i,t} = F_i\left(V_{k_1,\tau_{k_1}}, V_{k_2,\tau_{k_2}}, \ldots, V_{k_n,\tau_{k_n}}\right)$$

where $1 \leq i, k_1, \ldots, k_n \leq N$ are the node indices,$\tau_{k_1}, \ldots, \tau_{k_n} \leq t$ denote the time step when the state of nodes $k_1, \ldots, k_n$ was last updated. We will revisit the details of time implementation in section 3.3. As time evolves, the system state (N-dimensional vector) traverses the state space and after a finite duration transient behavior settles into an attractor (stable dynamic behavior). Two types of attractors are possible: i) after hitting a certain state, any future updating results in the same state, thus the system reaches a 'fixed point' or steady state, or ii) there exists a certain small set of states of the system *G* which the system keeps revisiting; that is, any updating of the system state will carry it to one of the states that belongs *G*. Depending on how updating is implemented, the second type of attractor of the system (non-steady state) can take on two possible forms, either an oscillation (a series of system states that repeat regularly) or a loose attractor (a random sequence of system states that is generated from a finite pool of states). Please see section 2.3.5 for further details.

Boolean dynamic modeling of a biological system is comprised of the following steps: reconstruct the network based on biological knowledge; determine the Boolean transfer functions; choose an updating scheme; determine the initial state of the model; analyze the model including its attractors and state space; validate the model (reproduction of known results) and finally, study novel scenarios, e.g. robustness against disruptions, make useful predictions and inferences. We next look at each step in more detail and elaborate them through examples.

2.3.1 Reconstruct the network based on biological knowledge

The first step of Boolean dynamic modeling is to represent the system under study by a network, denoting the relevant elements of system by nodes and their pairwise relationships by

edges as discussed in section 2. This is accomplished through extensive literature search and compilation. Experimental databases available online, such as Transcription Factor Database (TRANSFAC, 32) and Kyoto Encyclopedia of Genes and Genomes (KEGG, 33), can be used for data mining to deduce causal relationships between components. Many, but not all, experiments indicate direct interactions or regulatory relationships of elements, such as transcription factor-gene interactions, enzymatic activities and protein-protein interactions. Genetic knockout or pharmaceutical evidence such as exogenous application of a certain chemical indicates regulatory relationships indirectly. With such relationships further inference and interpretation of experimental results might be needed to obtain the most proper regulatory relations to represent in the network (10). For instance, experimental identification of the change in the activity level of a protein after a certain stimulus was applied implicates the protein as a potential downstream responder to that stimulus. Similarly, if the over-expression of a gene results in the downregulation of another gene or abnormally low activity of an enzyme, this implies that the gene negatively affects its target. A change in the concentration of a protein following a genetic knockout or over-expression can indicate the protein of interest as a potential regulation target of the gene being manipulated.

One can summarize the biological facts collected and list them in a table in order to synthesize and to present in a clear way all knowledge ready to be represented by a graph. We present an example consisting of five nodes in Table 2-1.

Biological Evidences
Input activates both A and C.
A activates B, but is inhibited by B at the same time.

	C is inhibited by A.
	Output is activated by C.

Table 2-1. A listing of biological evidences describing the causal relationships between components of a system.

The network constructed based on Table 2-1 is shown in Figure 2-2.

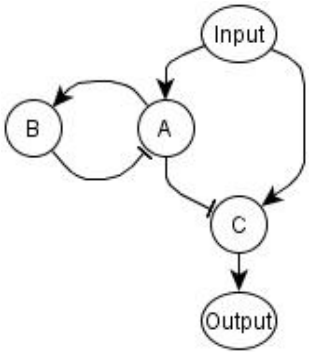


Figure 2-2. The graph representation of the example system described in Table 1. Input is the signal to the whole system and the sole source node (node with no incoming edges), and Output is the only sink node (node with no outgoing edges). '→' represents activation; '—|' represents inhibition.

Apart from straightforward evidences that indicate the regulatory relationship between two components as in Table 1, often experimental results can lead to complex inferences such as "A promotes the process through which B activates C", the simplest case of which is that A catalyzes the reaction from B to C, or "B induces the synthesis of C only in the absence of A". Representation of such cases necessitates groups of three or more directed edges combining multiple nodes (as in Figure1C).

More often than not, even after careful synthesis the knowledge about a biological system is not sufficiently complete. Let's consider the hypothetical scenario of a signal that is known to

function as an inhibitor of a certain output. Previous experiments indicate that the signal is activating node A, and the output is negatively regulated by node B, but no assay of the interaction between A and B has been carried out. Under such circumstances, one essentially needs to make reasonable and parsimonious assumptions to bridge the gap between existing evidences. In this case, a positive mass flow or regulatory relation oriented from A to B completes a feasible signal transduction pathway which is consistent with prior knowledge. In case where contradicting evidences are presented, one needs to critically examine and compare the methods by which the results are obtained, the environments under which the experiments were performed, and accept the better supported relation.

Manual assembly and interpretation can become daunting for systems containing hundreds of nodes and abundant causal relationships. The methodology (34) has been developed software packages (e.g. NET-SYNTHESIS, http://www.cs.uic.edu/~dasgupta/network-synthesis, 35) aimed specifically at tackling the problem of network assembly at large scale. Taking a text file describing all causal relationships between system components as input, the software synthesizes and generates a network representation of the system, and outputs a file containing information of all the edges.

2.3.2 Determine the Boolean transfer functions

After the network backbone is assembled, the second crucial step towards dynamic simulation is to determine the Boolean transfer functions that govern the state transition of nodes through time. A node *i* might have one or more upstream regulators in the network. The Boolean transfer function expresses the way the states of these regulators are combined through the Boolean operators *AND*, *OR*, and *NOT*. The transfer functions are also referred to as Boolean rules. For clarity in the following examples we will denote the state of nodes by the node name and simplify

the representation of time by only considering current and future time, the latter denoted by using an asterisk on the node name. When there is a single input, the state of the output at the next time step can take on one of the two possible states available, namely

$Output^{*}$=Input,

if there is a positive relationship between the output and the input, or

$Output^{*}$=*NOT* Input,

if the input suppresses the output. When there are two or more input nodes, their states need to be combined via Boolean functions in a way that is consistent with available knowledge. A biochemical synthesis reaction A+B→C can naturally be represented by the Boolean rule: C^{*} = A *AND* B. The statement "B induces the synthesis of C only in the absence of A" can be represented by the Boolean rule C^{*} = *NOT* A *AND* B, since both conditions: stimulus B as well as the absence of inhibitory factor A need to be satisfied in order for C to be synthesized. On the contrary, if for instance two inputs function largely independently, do not exhibit synergy and can substitute one another, then the Boolean *OR* function should be used, e.g. C^{*} = A *OR* B. The information of dependencies of input nodes can be obtained through literature search.

If in the system shown on Figure 2-2 both the presence of Input and the absence of Bare required for the activation of A, then the Boolean function of A can hence be written as:

A* = Input *AND* (*NOT* B)

Similarly, if either the absence of A or the presence of Input is sufficient to activate C, this leads to:

C* = (*NOT* A) *OR* Input

The Boolean transfer functions of the system on Figure 2-2 are summarized in Table 2-2:

Node	Boolean Rule (Transfer Function)
Input	—
A	A* = Input *AND* (*NOT* B)
B	B* = A
C	C* = (*NOT* A) *OR* Input
Output	Output* = C

Table 2-2. Boolean rules (transfer functions) of the system shown in Figure 2-2, for specific assumptions on the combination of multiple inputs. These rules govern the state transitions of the nodes in the system.

The more inputs a node has, the more possible ways there are to combine all the input states. One needs to be particularly careful in such cases in order to obtain the rule that is able to generate dynamic simulation results that best fit the existing experimental data. It is likely that multiple "trial and error" processes will have to take place before an optimal solution is found.

Instead of describing the state transition relationship with compact Boolean rules, one can also list all possible combinations of input states of a certain node *i* and assign output states to each of such combinations. This is called the truth table of a certain Boolean rule, which is equivalent to the rule itself. If node *i* has *m* input nodes, then the truth table will have 2^m rows.

2.3.3 Updating schemes and incorporating time

We have now explained the network structure and the Boolean transfer functions that govern the state transitions of nodes ready. These two pieces comprise the major steps of the model. The next step, discretization of the continuous time stream into steps, is also nontrivial. One has

different choices in terms of the number of nodes to be updated at each time step, and the order in which the updating of those nodes is to be carried out, and these choices may affect the results.

Assume that the total number of nodes in the network is N. Starting from the general expression of Boolean rules established based on biological knowledge, we will look at each updating scheme in detail.

Suppose a node *i* has *n* input nodes and the Boolean transfer function of node *i* can be written as follows:

$$V_{i,t} = F_i\left(V_{k_1,\tau_{k_1}}, V_{k_2,\tau_{k_2}}, \dots, V_{k_n,\tau_{k_n}}\right)$$

where $V_{i,t}$ is the state of node i at time step t, $1 \le i, k_1, \dots, k_n \le N$ are the node indices, $\tau_{k_1}, \dots, \tau_{k_n} \le t$ denote the time step of the last update of the state of nodes $k_1, \dots, k_n$. Two major types of updating exist, namely synchronous update (36) and asynchronous update (7).

For synchronous updating every node is updated once at each time step, using the states of the input nodes at earlier time steps as inputs to the transfer functions. In other words, $\tau_{k_j} \le t - 1$, where $1 \le j \le n$. The fact that the input states to each and every transfer function are from earlier time steps makes the order in which all nodes are updated at each time step irrelevant, since all inputs have already been fixed before the updating.

Synchronous updating has a simple and straightforward formalism and leads to reproducible state changes. It, however, overlooks the differences of time scales on which biological processes are taking place, which can range wildly from milliseconds for protein phosphorylation and posttranslational activities to hundreds of seconds for transcription and transcriptional regulations (37). Asynchronous updating (7), which provides more detailed

tracking of time scales and temporal orders (10-12), is devised in order to account for the diversity in lengths of biological processes.

We will cover three types of asynchronous updating: i) general asynchronous updating (38), ii) random order asynchronous updating (38, 39), and iii) deterministic asynchronous updating (40).

In general asynchronous updating (38), only one node is randomly selected and updated at each time step. The states of the inputs to the node that is being updated are from earlier time steps, which means $\tau_{k_j} \leq t-1$ for $1 \leq j \leq n$.

In random order asynchronous updating (38, 39), at each time step, a random permutation of the node labels (from 1 to N) is first generated, and the updating is carried out in that sequence in that time step. The random permutation of 1 to N is regenerated at each time step. Consider node *i* with n inputs and its updating:

$$V_{i,t} = F_i\left(V_{k_1,\tau_{k_1}}, V_{k_2,\tau_{k_2}}, \ldots, V_{k_n,\tau_{k_n}}\right)$$

If node k_j comes before node *i* in the random sequence, meaning that by the time node *i* is being updated, node k_j has already been updated at the current time step to its latest state, $\tau_{k_j} = t$ should be used. Otherwise, if node k_j comes after node *i* in the random sequence, meaning that at the time node *i* is being updated, the state of node k_j has not been touched yet at the current time step, $\tau_{k_j} \leq t-1$ should be used.

In deterministic asynchronous updating (40), every node is associated with an intrinsic time scale γ_i. The updating of node *i* only takes place if the current time t is a multiple of γ_i. One can clearly see that γ_i is the effective 'pace' of each reaction. In other words,

$$V_{i,t} = F_i\left(V_{k_1,\tau_{k_1}}, V_{k_2,\tau_{k_2}}, \ldots, V_{k_n,\tau_{k_n}}\right), \;\; t = c\gamma_i, c \; is \; a \; positive \; integer.$$

When the updating of node *i* takes place, the states of its input nodes should be taken from the latest available time step, namely $\tau_{k_1}, \ldots, \tau_{k_n} \leq t - 1$.

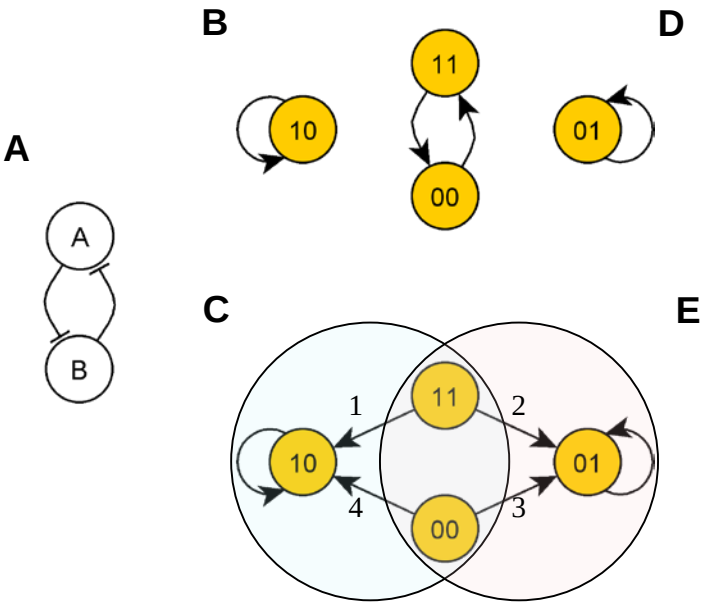
Several software packages can be used to carry out dynamic simulations based on Boolean network modeling on biological systems. An open-source Python package, BooleanNet (41), is available online at http://code.google.com/p/booleannet/. The input to the program is a text file containing all the Boolean transfer functions, and all of the updating schemes can be selected and readily implemented. Other software packages include Cell Net Analyzer (MatLab package, 42), GINsim (43), and BoolNet (R package, 44).

2.3.4 Network initialization

The system state is represented by an N-dimensional vector, for a system of N nodes (elements) (see the beginning of section 3). To perform dynamic simulations of the system under the Boolean network framework, one needs to specify the first state in the state sequence (trajectory), namely the initial state of the system. The initial state of each node is determined such that it is consistent with known biological facts or experimental evidence. If an intracellular substance is known to be present under all conditions, it can be initialized to be in the 1, or *ON* state. Initial states can also be assigned based on the question of interest. For example, one can implement a gene knockout by the initial and sustained *OFF* state of that particular gene. If there is insufficient experimental information concerning the concentration or activity level of an

element, one can also randomize the initial state of that node. Therefore, a large number of runs of dynamic simulations can be carried out with each run randomizing the initial states of the nodes that cannot be predetermined (10). The system will sample different potentially viable initial states and henceforth take varying routes in the state space.

Starting from the initial state, the system will evolve as time progresses and should eventually settle down into an attractor, that is, a time-invariant steady state or an (ordered or unordered) repetition of a certain finite set of states. These attractor states have been proposed as representations of cell fates going back to the 1940's work of C. Waddington (45, 46) and later by S. Kauffman (6). The complete set of states that can potentially reach a certain attractor through an updating scheme forms the basin of attraction of *that* attractor. Since synchronous updating is deterministic, the basins of attraction for different attractors will be distinct, whereas under asynchronous updating, the basins of attraction for different attractors could share states and be partly overlapping with each other as illustrated by the following example in Figure 2-3 (39).



Node	Boolean Rules
A	A* = *NOT* B
B	B* = *NOT* A

Edge	Node updating order
1, 3	B, A
2, 4	A, B

Figure 2-3. Ilustration of the basins of attraction of a system under different updating schemes. The system is composed of two mutually inhibiting nodes A and B (A). The Boolean functions are given in table (D). (B) and (C) show the state transition graphs of the system under synchronous update (B), and random order asynchronous update (C). The 2-digit vectors inside the circles represent the states of node A and B from left to right. In the case of random order update (C) the correspondence between the order in which A and B are updated and the particular state transitions indicated by edge labels is given in table (E). E.g. starting from an initial state of 11, if A is updated first and B is updated second, state 11 will evolve into state 01. The two unlabeled self-loops in (B) and (C) indicate that if the system is currently in one of the two states, the future state will be the same as the current state, regardless of the manner of update, thus these are fixed points of the system. Under synchronous update (B), the system has 3 attractors: 2 fixed points: 01 and 10, and a cyclic attractor. Only the fixed points are preserved under asynchronous update (C). In (C), the states 11, 00, 10 form the basin of attraction of attractor 10, and states 11, 00, 01 form the basin of attraction of attractor 01. The states 11 and 00 belong to both basins of attraction.

2.3.5 Steady state analysis

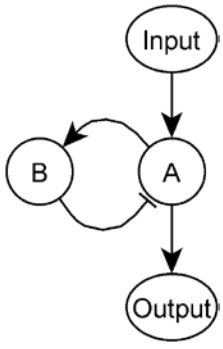
The system will have the same steady states (fixed points) under both synchronous and asynchronous updating, due to the fact that a steady state repeats itself infinitely, making the order in which the nodes are updated irrelevant. BooleanNet (41) provides functions to detect the steady states of the system. Cell Net Analyzer (42) is also able to probe the steady states of the system, which are called ‘logical steady states’ in the package.

Before we illustrate in a step-by-step manner how to determine the steady states of a Boolean network, we note that there are two main focuses encountered in dynamic analysis of biological systems: 1) determining the attractors of the whole system (of all nodes), and 2)

determining the attractors of a small set of designated *output nodes* of the Boolean network. The second focus is necessarily a subset of the first. The output-oriented analysis is oftentimes not only simpler, but also more relevant to signal transduction pathway-related research, where no more than several inputs (signals) are considered and the system's response is usually characterized by a single output node. In modeling gene regulatory networks, usually the attractors and the dynamic sequence of the whole system correspond to known biological events such as certain phases of the cell cycle (8, 47), apoptosis and cell differentiation (48), and thus the first focus, identifying attractors of the whole system, is most relevant.

The fixed points of a Boolean network model can be determined in several ways. One can analytically solve the Boolean equations. Since the system being in the steady state means that the state vector remains time-invariant, the future state of any node will be the same as its current state. Thus, the time-dependent features (indices or asterisk marks) of the Boolean equations can be removed, and the set of resulting time-independent equations can readily be solved. One can also do repeated dynamic simulations of the system, updating the nodes' states according to their Boolean rules. One can also draw useful conclusions from the existence of particular interaction patterns (called 'network motifs') (49), such as feed-forward or feed-back loops (as first suggested by R. Thomas, 50). Let's look at the following examples:

A



B

Node	Boolean Rule
Input	—
A	$A^* =$ Input *AND* (*NOT* B)
B	$B^* = A$
Output	$Output^* = A$

Figure 2-4. A four-node signal transduction network in which A is activated by Input, and its activation also *requires* the absence of B. (A) Network representation; (B) the list of Boolean rules.

In the example of Figure 2-4, the system (thus also the Output) has a single attractor for any update method. For synchronous update, the attractor is a cycle of period 4 (Figure 2-5A). Under general asynchronous updating the attractor spans the whole state space. Every state in the latter attractor has multiple edges, both incoming and outgoing (Figure 2-5B). This is because a state can have different successor states if a different node is updated. The maximum number of possible outcomes from a state under the general asynchronous scheme equals the total number of nodes in the system. The system state will traverse all states in the attractor shown in Figure 2-5B), following the outgoing edges with a certain calculable probability. For both types of update, the average state level of the node Output (the last digit) in the attractor is 0.5. This could be interpreted as a 50% up time of a certain system activity.

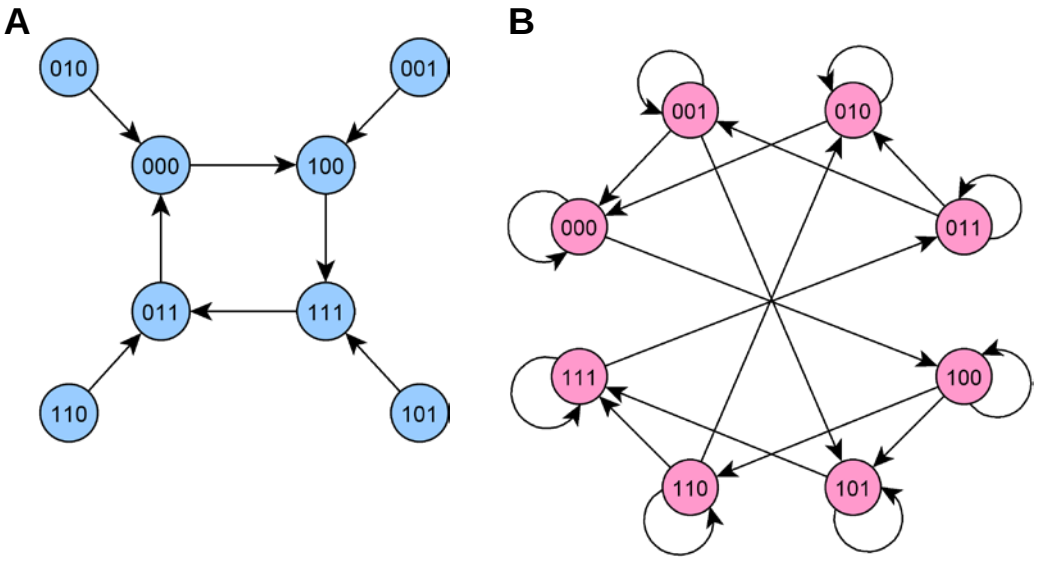
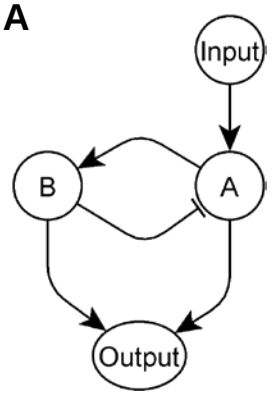


Figure 2-5. The state transition graph of the system described in Figure 2-4. The 3-digit vectors inside the circles represent the state of node A, B, and Output from left to right. The state of the Input is fixed at 1, indicating a constant ON signal to the system (not shown in the graph). A self-loop on a certain state indicates that if one of the nodes is updated, the system state remains the same. For 3 nodes, a total of 8 states exist. (A) Synchronous update. All 8 states are located in the same basin of attraction, of which 4 form the attractor. (B) General asynchronous update. The attractor is formed by all 8 states. These figures were generated by BooleanNet.

The critical network motif (49) of the original network is the negative feedback on node A. It is the negative feedback, coupled with the Boolean *AND* rule for node A, which generates the oscillatory behavior of the system.

A



B

Node	Boolean Rule
Input	—
A	$A^* = \text{Input } AND \text{ } (NOT \text{ B})$
B	$B^* = A$
Output	$\text{Output}^* = A \text{ } OR \text{ B}$

Figure 2-6. A four-node signal transduction network wherein A is activated by Input, and its activation also *requires* the absence of B. Either A or B is sufficient to activate the Output. (A) The network representation; (B) the list of Boolean rules.

The next example network (Figure 2-6) differs from the previous one (Figure 2-4) in that it contains one more edge from B to the Output, and the state of the Output is determined by A and B independently, in a Boolean *OR* relation. We already know that the negative feedback on node A will generate an oscillatory behavior for itself as well as for node B, but does the new edge help

improve the activation of the Output? The answer is yes. Under both updating schemes the average state level of the Output is ~0.75. If the Boolean relation between A and B in the Output function is changed to *AND*, then the Output level drops to ~0.25. It is indeed reasonable that a more stringent condition for activation will decrease the average Output level. Under synchronous update, the Output (and also the system) will have period-4 oscillations. In the state transition graphs of the system (Figure 2-7), (a) corresponds to Output* = A *OR* B, and (b) to Output* = A *AND* B.

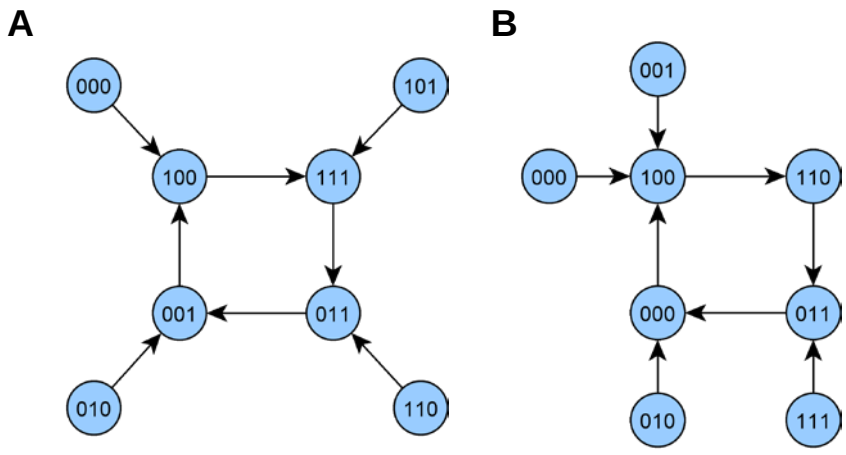
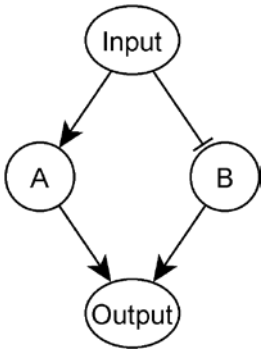


Figure 2-7. The state transition graphs corresponding to the network in Figure 4 with two variants of the Boolean transfer functions. The 3-digit vectors represent the state of node A, B, and Output from left to right. The state of the Input is fixed at 1. (A) Output* = A *OR* B. An average level of 0.75 of the node Output is observed in the attractor, which is a period-4 oscillation. (B) Output* = A *AND* B. An average level of 0.25 of the node Output is observed in the attractor.

A **B**



Node	Boolean Rule
Input	—
A	A* = Input
B	B* = *NOT* Input
Output	Output* = A *AND* B

Figure 2-8. A four-node signal transduction network in which A is activated by Input, B is inhibited by Input, the activation of the Output requires the presence of both A and B. (A) The network representation; (B) the list of Boolean rules.

Finally, let's look at yet another example (Figure 2-8). The characteristic feature of this network is two paths that both connect the Input and the Output, but functionally contradict each other. Given the *AND* rule for the Output, the steady state of Output is *OFF*, and more importantly, this outcome is independent from the state of the Input: the Output is decoupled from the Input in this case. If *OR* is used in the transfer function of the Output, its state will be at constant *ON* instead, but it will still be independent from the Input state.

Now let's apply the dynamic analysis to the network in Figure 2-2 and see whether a steady state of the Output is possible. Given that the Input node is *ON*, the negative feedback on node A will generate an oscillation of states for node A as well as node B. However, since the Boolean rule for node C is C* = (*NOT* A) *OR* Input, the oscillation coming from branch A will be combined by an *OR* rule with the *ON* state of the Input. The Output node will henceforth reach an *ON* steady state. The attractor for the state vector of the system, on the other hand, will be a limit cycle, since a subset of the nodes is oscillating.

In general, multiple network motifs can be present in the graph representation of a system under study, and each has its own contribution to generate advantageous dynamic behaviors, stabilize the system function, prevent extreme behavior, or provide system redundancy and consequently robustness (51-53). Last but not least, the conditionality present in the Boolean transfer functions is often critical in determining the system behavior.

2.3.6 Model validation: reproduction of known results

An important step toward obtaining a successful model is to examine the dynamic sequence of system states in detail, and compare with biological evidence. If, for instance, under reasonable assumptions, starting from a viable initial state, an oscillation of the output is observed whereas it is biologically known that the output should exhibit a constant *OFF* state under such conditions, further modification of one or more Boolean transfer functions or even the underlying network structure is called for. Such discrepancies can also be a potential indication of components missing from the model. One needs to scrutinize all possibilities and make the most accurate, biologically truthful model revision by changing Boolean rules, rewiring edges between nodes, and/or adding/removing nodes to/from the network, so as to reproduce as many known results as possible. In a noteworthy example, Samaga *et al*. performed a comparison between simulation results generated by a Boolean logic model constructed based on the literature and data collected from cells on ErbB receptor phosphoprotein signaling (54). They came to a set of 11 hypotheses regarding ErbB signaling resulting from the discrepancies between simulation results and experimental observations, among which 5 were supported by literature, 5 led to further modifications of the model, and 1 implied the absence of specificity expected from a small molecular inhibitor.

2.3.7 Robustness against disruptions and useful implications

An important additional assessment is whether the model is robust in terms of changes in interactions or Boolean transfer functions. Models that are extremely fragile to such changes may not be a good representation to biological systems, as the real systems exhibit substantial robustness to changes in concentrations, reaction rates or even mutations (55-59). As currently comprehensive models of signal transduction systems are rare, the model should show reasonable robustness to changes in the network structure to instill confidence that its results will still stand after new components or interactions are discovered. The ability of the model to maintain the wild-type response under small topological perturbations can be tested by adding or deleting a randomly selected node or edge, rewiring edges in the network randomly (for example changing any pair of parallel edges to cross-edges) or make an inhibitory interaction into activation or *vice versa*. For example, if in Figure 2-1 node B is knocked out (constitutively *OFF*), the Output of the system will *not* be affected, due to the *OR* relation in the Boolean rule of node C. However, if the edge from Input to C becomes nonfunctional, a new oscillatory behavior of the system is exhibited. It is also possible that existing attractor basins get altered, or completely eliminated, and new ones arise from such perturbations of the system.

Validated models can be used to predict the outcomes of "what if" scenarios, cases that were not yet studied experimentally, and they can generate testable predictions and significant insights. For example, Mendoza formulated a logic-based model of interactions among cytokines and transcription factors in helper T (Th) cells (28). Dynamic simulation under all combinations of initial node states revealed four steady states: one corresponding to naïve Th cells, one corresponding to Th2 cells, and two corresponding to Th1 cells. The two Th1 cell attractors indicated two Th1 cell subpopulations with different levels of IFNγ secretion, nevertheless the level of the IFNγ receptor was the same in both attractors, a result supported by the literature.

Mendoza studied in detail how node perturbations (knockout or overexpression) change the differentiation fate of Th cells. Several of the model results were supported by the experimental literature data while numerous others are novel predictions (28).

We next take a closer look at Boolean models applied in two different contexts: survival signaling in T cells in T-LGL leukemia (12), and interactions among pathogens and a mammalian immune system (11). The second example also demonstrates that Boolean modeling is a general approach that can be applied at different levels of biological organization, from the molecular to the population level (60).

2.4 Application Examples

2.4.1 T-LGL leukemia network modeling

Cytotoxic T lymphocytes (CTLs) are a type of T cells that are activated in order to clear virus-infected cells. The activated population undergoes a subsequent activation-induced cell death and eventually reaches a balance between proliferation, survival, and apoptosis. T cell large granular lymphocyte (T-LGL) leukemia is characterized by an abnormal clonal proliferation of cytotoxic T lymphocytes that escaped activation-induced cell death (12). The diseased T cells are insensitive to Fas-induced apoptosis and henceforth remain long-term competent. Zhang *et al*. constructed a T-LGL survival signaling network and a Boolean model of the network. Through a vast literature search and the use of NET-SYNTHESIS (35) to simplify the original network, a final network of 58 nodes and more than 100 edges was constructed. The input node to the system is ‘stimuli’, representing antigen stimulus and the output node of the system is ‘apoptosis’, representing activation-induced cell death, which is the outcome of the system under normal conditions. The model (12) identified two proteins—IL-15 and PDGF— which are critical in

inducing and maintaining the diseased cell behavior. The overexpression of the two is a sufficient condition to reproduce all known leukemic abnormalities. Furthermore, key survival mediators, such as NFκB, SPHK-1 and GAP were also proposed. These nodes settle down into *ON* or *OFF* state in T-LGL leukemia, and if the state of at least one of them is flipped artificially (*ON*→*OFF* or *OFF*→*ON*), apoptosis is induced. This offered potential therapeutic targets for curing T-LGL leukemia. Several of these predictions were validated experimentally. The study also demonstrated that it is possible to integrate qualitative information regarding normal and diseased cellular signaling pathways into the same network, and discrete dynamic modeling of this network can be used to generate deep insight and validated predictions.

2.4.2 Pathogen-Immune response network

Interactions between an invading pathogen and the responses of the immune system form a complex signaling network (61-63). Pathogens seek to evade or disrupt the host immune response to ensure access to nutrients and self-proliferation. For example, bacteria persist within hosts by interfering with antigen production or processing, subverting phagocytosis by immune cells, and/or by promoting the anti-inflammatory responses of the host that under normal conditions deactivate the protective effectors of the host. Pathogens therefore serve as the input to such a network and the pathogen-host immune system interaction can result in either the clearance of the pathogen or persistent disease.

Thakar *et al.* (11) performed a comparative study between two species of the same genus *Bordetella* to better understand the different ways in which various virulence factors modulate immune responses and consequently the adaptability of the pathogen and survivability of the hosts (11, 63). Two separate network models for the interaction of these pathogens with their hosts were

synthesized. The nodes of the networks were comprised of immune cells, cytokines, antibodies as well as conceptual nodes such as phagocytosis, among which 18 were common to both networks. 'Bacteria' is the input to the network, and phagocytosis can be considered the output, which in turn leads to the clearance of Bacteria, generating a negative feedback edge from the output to the input. The edges in the network represent cytokine production, cell recruitment and differentiation, and intercellular signaling processes.

Each network was then translated into a predictive dynamic model, using Boolean transfer functions reconstructed from the literature, and was validated by experimental observations. The model led to an understanding of puzzling differences between the two pathogens, e.g. that antibody transfer at the time of infection (similar to immunization) aided clearance in *B. bronchiseptica* but not in *B. pertussis*. On the contrary, the model predicted that for both pathogens a secondary infection would be cleared earlier than a first-time infection. Follow-up experiments were performed where a *B. pertussis* or *B. bronchiseptica* second challenge was given to convalescent hosts. The secondary infection was indeed cleared in 15 days, which is significantly less than the duration of a first-time infection. The model also predicts that certain cytokines are an equally effective prophylactic measure against both pathogens. The study identified three phases in *Bordetella* infections and offered a methodology of *in silico* evaluation of putative medical treatments. This methodology generates novel insights into the interplay between pathogen virulence factors and host immune response system, and can be readily adapted to similar systems.

2.5 References

1. Lee, T. I., Rinaldi, N. J., Robert, F., Odom, D. T., Bar-Joseph, Z., Gerber, G. K., Hannett, N.

M., Harbison, C. T., Thompson, C. M., Simon, I., Zeitlinger, J., Jennings, E. G., Murray, H. L., Gordon, D. B., Ren, B., Wyrick, J. J., Tagne, J.-B., Volkert, T. L., Fraenkel, E., Gifford, D. K., and Young, R. A. (2002). Transcriptional regulatory networks in *Saccharomyces cerevisiae*. *Science* 298, 799–804.

2. Schwikowski, B., Uetz, P., Fields, S. (2000). A network of protein–protein interactions in yeast. *Nature Biotechnology* 18, 1257–1261.
3. Stelzl, U., Worm, U., Lalowski, M., Haenig, C., Brembeck, F. H., Goehler, H., Stroedicke, M., Zenkner, M., Schoenherr, A., Koeppen, S., Tim, J., Mintzlaff, S., Abraham, C., Bock, N., Kietzmann, S., Goedde, A., Toksoz, E., Droege, A., Krobitsch, S., Korn, B., Birchmeier, W., Lehrach, H., Wanker, E. E. (2005) A Human Protein-Protein Interaction Network: A Resource for Annotating the Proteome. *Cell* 122, 957–968.
4. Hatzimanikatis, V., Li, C., Ionita, J. A., Broadbelt, L. J. (2004). Metabolic networks: Enzyme function and metabolite structure. *Curr. Opin. Struct. Biol.* 14, 300–306.
5. Chen, K. C., Csikasz-Nagy, A., Gyorffy, B., Val, J., Novak, B., and Tyson, J. J. (2000) Kinetic analysis of a molecular model of the budding yeast cell cycle. *Mol. Biol. Cell* 11, 369–391.
6. Kauffman, S. A. (1969) Metabolic stability and epigenesis in randomly constructed genetic nets. *J. Theor Biol.* 22, 437–467.
7. Thomas, R. (1973) Boolean formalization of genetic control circuits. *J. Theor Biol.* 42(3), 563–585.
8. Li, F., Long, T., Lu, Y., Ouyang, Q. & Tang, C. (2004) The yeast cell-cycle network is robustly designed. *ProcNat AcadSci USA* 101, 4781–4786.
9. Davidich, M.I., and Bornholdt, S. (2008) Boolean network model predicts cell cycle sequence of fission yeast. *PLoS ONE* 3, e1672.
10. Li, S., Assmann, S. M., and Albert, R. (2006) Predicting Essential Components of Signal Transduction Networks: A Dynamic Model of Guard Cell Abscisic Acid Signaling. *PLoS Biol*. 4, e312.
11. Thakar, J., Pilione, M., Kirimanjeswara, G., Harvill, E. T., and Albert, R. (2007) Modeling Systems-Level Regulation of Host Immune Responses. *PLoSComput Biol*. 3, e109.
12. Zhang, R., Shah, M. V., Yang, J., Nyland, S. B., Liu, X., Yun, J. K., Albert, R., and Loughran, T. P., Jr. (2008) Network model of survival signaling in large granular lymphocyte leukemia. *ProcNatlAcadSci* USA 105, 16308–16313.

13. 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. *PLoSComput Biol*. 3, e163.
14. Chaouiya, C. (2007) Petri net modelling of biological networks. *Brief Bioinform*. 8, 210–219.
15. Peterson, J. L. (1981) Petri Net Theory and the Modeling of Systems, Prentice Hall PTR.
16. Sackmann, A., Heiner, M., and Koch, I. (2006) Application of Petri net based analysis techniques to signal transduction pathways. *BMC Bioinformatics* 7, 482.
17. Mendoza, L., Thieffry, D. & Alvarez-Buylla, E.R. (1999) Genetic control of flower morphogenesis in *Arabidopsis thaliana*: a logical analysis. *Bioinformatics* 15, 593–606.
18. Espinosa-Soto, C., Padilla-Longoria, P. & Alvarez-Buylla, E.R. (2004) A gene regulatory network model for cell-fate determination during *Arabidopsis thaliana* flower development that is robust and recovers experimental gene expression profiles. *Plant Cell* 16, 2923–2939.
19. Faure, A., Naldi, A., Lopez, F., Chaouiya, C., Ciliberto, A. &Thieffry, D. (2009) Modular logical modelling of the budding yeast cell cycle. *Molecular Biosystems* 5, 1787–1796.
20. Sanchez, L. &Thieffry, D. (2001) A logical analysis of the *Drosophila* gap-gene system. *J Theor Biol.* 211, 115–141.
21. Albert, R., and 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
22. Sanchez, L., and Thieffry, D. (2003) Segmenting the fly embryo: a logical analysis of the pair-rule cross-regulatory module. *J Theor Biol*. 224, 517–537.
23. Ghysen, A. & Thomas, R. (2003) The formation of sense organs in *Drosophila*: a logical approach. *Bioessays* 25, 802–807.
24. Gonzalez, A., Chaouiya, C., and Thieffry, D. (2008) Logical modelling of the role of the Hh pathway in the patterning of the *Drosophila* wing disc. *Bioinformatics* 24, i234–i240.
25. Diaz, J. & Alvarez-Buylla, E.R. (2006) A model of the ethylene signaling pathway and its gene response in *Arabidopsis thaliana*: pathway cross-talk and noise-filtering properties. *Chaos* 16, 023112.
26. Naldi, A., Carneiro, J., Chaouiya, C. &Thieffry, D. (2010) Diversity and plasticity of Th cell types predicted from regulatory network modelling. *PLoSComput Biol*. 6(9): e1000912.
27. Naldi, A., Berenguier, D., Faure, A., Lopez, F., Thieffry, D. &Chaouiya, C. (2009) Logical

modelling of regulatory networks with GINsim 2.3. *Biosystems* 97, 134–139.

28. Mendoza, L. (2006) A network model for the control and differentiation process in Th cells. *BioSystems* 84, 101–114.
29. Voit, E. O. (2000) Computational Analysis of Biochemical Systems—A Practical Guide for Biochemists and Molecular Biologists. *Cambridge University Press.*
30. Wang, R., and Albert, R. (2011) Elementary signaling modes predict the essentiality of signal transduction network components. *BMC Systems Biology* 5, 44.
31. Saez-Rodriguez, J., Alexopoulos, L. G., Epperlein, J., Samaga, R., Lauffenburger, D. A., Klamt, S., and Sorger, P. K. (2009) Discrete logic modeling as a means to link protein signaling networks with functional analysis of mammalian signal transduction. *Molecular Systems Biol.* 5, 331.
32. Wingender, E., Dietze, P., Karas, H., and Knuppel, R.(1996) Transfac: A database on transcription factors and their DNA binding sites. *Nucleic Acids Res* 24, 238–241.
33. Kanehisa, M., Goto, S. (2000) KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res*. 28, 27–30.
34. Albert, R., DasGupta, B., Dondi, R., Kachalo, S., Sontag, E., Zelikovsky, A., and Westbrooks, K. (2007) A novel method for signal transduction network inference from indirect experimental evidence. *J Comput Biol*. 14, 927–949.
35. Kachalo, S., Zhang, R., Sontag, E., Albert, R., and DasGupta, B. (2008) NET-SYNTHESIS: a software for synthesis, inference and simplification of signal transduction networks. *Bioinformatics* 24, 293–295.
36. Kauffman, S. (1993) Origins of order: self-organization and selection in evolution, Oxford Univ. Press, New York.
37. Papin, J. A., Hunter, T., Palsson, B. O., and Subramanian, S. (2005) Reconstruction of cellular signalling networks and analysis of their properties. *Nat Rev Mol Cell Biol*. 6, 99–111.
38. Harvey, I., and Bossomaier, T. (1997) Time out of joint: Attractors in asynchronous random Boolean networks. In ‘‘Proceedings of the Fourth European Conference on Artificial Life (ECAL97),’’ (P. Husbands and I. Harvey, eds.), pp. 67–75. MIT Press, Cambridge, MA.
39. Chaves, M., Albert, R., and Sontag, E. D. (2005) Robustness and fragility of Boolean models for genetic regulatory networks. *J Theor Biol*. 235, 431–449.
40. Chaves, M., Sontag, E. D., and Albert, R. (2006) Structure and timescale analysis in genetic

regulatory networks. *SystBiol (Stevenage)* 153, 154–167.

41. Albert, I., Thakar, J., Li, S., Zhang, R., and Albert, R. (2008) Boolean network simulations for life scientists. *Source Code Biol Med* 3, 16.
42. Klamt, S., Saez-Rodriguez, J., Gilles, E. D. (2007) Structural and functional analysis of cellular networks with CellNetAnalyzer. *BMC Systems Biology* 1, 2.
43. Gonzalez, A. G.,Naldi, A., Sánchez, L., Thieffry, D., Chaouiya, C. (2006) GINsim : a software suite for the qualitative modeling, simulation and analysis of regulatory networks. *Biosystems* 84(2), 91–100.
44. Mussel, C., Hopfensitz, M., and Kestler, H. A. (2010) BoolNet—an R package for generation, reconstruction and analysis of Boolean networks. *Bioinformatics* 26, 1378–1380.
45. Waddington, C. H. (1942) Canalization of development and the inheritance of acquired characters. *Nature* 150, 563–565.
46. Waddington, C. H. (1953) Epigenetics and evolution. *Symp. Soc. Exp. Biol* 7, 186–199.
47. Tyson, J., and Novak, B. (2001) Regulation of the Eukaryotic Cell Cycle: Molecular Antagonism, Hysteresis, and Irreversible Transitions. *J Theor Biol.* 210, 249–263.
48. Huang, A. C., Hu, L., Kauffman, S. A., Zhang, W., and Shmulevich, I. (2009) Using cell fate attractors to uncover transcriptional regulation of HL60 neutrophil differentiation. *BMC Syst. Biol.* 3, 20.
49. Shen-Orr, S., Milo, R., Mangan, S., and Alon, U. (2002) Network motifs in the transcriptional regulation network of *Escherichia coli*. *Nature Genetics*, 31, 64–68.
50. Thomas, R., D'Ari, R. (1990) "Biological Feedback." 1st ed. CRC Press, Florida.
51. Shinar, G., Milo, R., MR, M., and Alon, U. (2007) Input output robustness in simple bacterial signaling systems. *ProcNatlAcadSci USA.* 104(50), 19931–19935.
52. Tyson, J.J., Borisuk, M.T., Chen, K., and Novak, B. (2000) Analysis of complex dynamics in cell cycle regulation, in Computational Modeling of Genetic and Biochemical Networks, Cambridge, MA, MIT Press. (J.M. Bower, and H. Bolouri, Eds), pp. 287–305
53. Bhalla, US, and Iyengar, R. (2001) Robustness of the bistable behavior of a biological signaling feedback loop. *Chaos*. 11(1), 221–226.
54. Samaga, R., Saez-Rodriguez, J., Alexopoulos, L. G., Sorger, P. K., and Klamt, S. (2009) The Logic of EGFR/ErbB Signaling: Theoretical Properties and Analysis of High-Throughput Data. *PLoSComput. Biol.* 5, e1000438.

55. vonDassow, G., Meir, E., Munro, E.M. & Odell, G.M. (2000) The segment polarity network is a robust developmental module. *Nature* 406, 188–192

56. Alon, U., Surette, M.G., Barkai, N. &Leibler, S. (1999) Robustness in bacterial chemotaxis. *Nature* 397, 168–171.

57. Eldar, A., Dorfman, R., Weiss, D., Ashe, H., Shilo, B.Z. &Barkai, N. (2002) Robustness of the BMP morphogen gradient in *Drosophila* embryonic patterning. *Nature* 419, 304–308.

58. Csete, M. & Doyle, J. (2004) Bow ties, metabolism and disease. *Trends in Biotechnology* 22, 446–450.

59. Conant, G.C. & Wagner, A. (2004) Duplicate genes and robustness to transient gene knock-downs in Caenorhabditiselegans. *Proc R SocLond B BiolSci* 271, 89–96.

60. Campbell, C., Yang, S., Albert, R., and Shea, K. (2011) A network model for plant–pollinator community assembly. *PNAS* 108, 197–202

61. Janeway, C. A. Jr., Travers, P., Walport, M., Shlomchik, M. J. (2001) Immunobiology. The Immune System in Health and Disease. 5th edition. *New York: Garland Science.*

62. Mills, K. H., Boyd, A. P. (2007) Topley and Wilson's microbiology and microbial infections. 10th edition. In: Kaufmann SH, Steward M, editors. *London: Edward Arnold*.

63. Thakar, J., and Albert, R. (2010) Boolean models of within-host immune interactions. *Current Opinion in Microbiology.* 13, 377–381.

64. Glass, L. (1975). Classification of biological networks by their qualitative dynamics. *J. Theor. Biol.* 54, 85–107.

Chapter 3

Multi-level Modeling of Light-Induced Stomatal Opening Offers New Insights into its Regulation by Drought

This chapter follows closely my work that has been published in *PLoS Computational Biology* **10**, e1003930. I performed all the model-related aspects of the project, including the model construction, validation and dynamic simulations.

3.1 Introduction

Stomata are small pores located in the epidermes of plants that allow carbon dioxide (CO_2) uptake for photosynthesis as well as diffusion of O_2, produced by photosynthetic reactions, from the plant to the atmosphere. They are also the sites of water vapor loss through transpiration. Stomata are bordered by pairs of guard cells, the swelling of which leads to stomatal opening (enlargement of the pore), while their shrinking leads to stomatal closure. The size and shape change of the guard cells is due to their uptake or loss of water, which is driven by changes in cellular osmotic potential as a result of the accumulation or depletion of solutes. Guard cells are sensitive to multiple external and internal stimuli, e.g. light, intercellular CO_2 concentration (C_i), the stress hormone abscisic acid (ABA), and vapor pressure difference (VPD) between the leaf interior and the surrounding atmosphere [1-9]. Guard cells have photoreceptors for red and blue light, and guard cell responses to light of these wavelengths are the main focus of this work. As stomatal aperture regulation has a major impact on both the hydration status and the photosynthetic status of the plant, guard cells' sensitivity to stimuli is vital to the survival of vascular terrestrial plants. Plants' successful adaptation to the environment influences all life-forms on Earth. In particular, better understanding of the signaling and regulatory networks involved in stomatal responses is a necessary step toward improving the drought tolerance of crops.

Guard cells have long been a popular system for dissecting the functions of individual genes and proteins within signaling cascades. The most studied signals are blue light and ABA [10]. There has been extensive experimentation carried out to elucidate the roles of key signaling proteins, enzymes, and small molecules in these signal transduction pathways, and to identify the relationships between diverse components in the system. Numerous experiments have addressed the roles of light quality [3,5,6], C_i [11], and VPD [5,12]. A synergistic action between blue and red light in the formation of malate, a major intracellular osmoticum, was discovered [13]. Phototropins were identified as blue light-specific photoreceptors of guard cells [14-16], mediating blue light-specific stomatal opening. New evidence constantly adds to our knowledge on guard cell functioning, e.g. the recently discovered inhibition by phosphatidic acid (PA) of blue light-induced stomatal opening via type 1 protein phosphatase (PP1) [17] and the relationship between the activation of the H^+-ATPase and light quality [18]. Much less has been done, however, on a systems level to synthesize all existing evidence into a network model of light-regulated stomatal opening, or to elucidate the crosstalk between different signal transduction cascades, such as those triggered by light and ABA. One such pioneering work was done by Li *et al.* on modeling the ABA signal transduction network leading to stomatal closure [19]. That work synthesized the published evidence for direct interactions and indirect causal effects between cellular components into a consistent network of ABA-induced closure and formulated a Boolean dynamic model that recapitulated or predicted a large number of knockout phenotypes. Another recent systems level advance is the development of the OnGuard software that incorporates ion transporters at the guard cell plasma and vacuolar membrane, the salient features of osmolyte metabolism, and the major controls of cytosolic Ca^{2+} concentration and pH [20]. In this software, and models that use it [21,22], the light signal transduction pathways are approximated by a pre-defined, light-dependent

increase in the activities of all ion-translocating ATPases at the plasma and vacuolar membrane, and in sucrose and malate synthesis. That work does not consider light of different wavelengths nor the specific mechanisms through which the different types of light signals are perceived and transduced.

Given the abundance of experimental results regarding stomatal opening and its regulation, dynamic modeling of the full light-stimulated stomatal opening process and its inhibition by ABA is now tenable, and is the focus of this work. I synthesized more than 85 articles describing experimental observations into a comprehensive network of 70 components, of which 4 are signals (blue light, red light, CO_2, and ABA), and stomatal opening is the sole output. The network incorporates in a parsimonious manner more than 150 interactions or causal relationships between components. I developed a dynamic model based on the network by characterizing each component with discrete activity levels and by describing its regulation with a combination of logic and algebraic functions. The multiple activity levels of the components and the detailed updating functions offer a biologically more accurate representation of the system than Boolean models; for example, the output node, stomatal opening, has more than 20 levels in the model, ranging from 0 to 14. The model has a repertoire of more than 10^{31} distinct states, which gives it substantial dynamic richness and makes it one of the most complex dynamic models of biological systems (see also [23-26]). At the same time the discreteness of the states maintains the computational simplicity of the model. The model recapitulates a comprehensive array of known behaviors and phenotypes. Since the model is made up of node-level information (i.e. the regulatory function of each component), this agreement serves as validation. The model enables an unprecedented understanding of the regulation of stomatal opening and predicts new phenotypes caused by the disruption of components. Moreover, the model reveals aspects of the

system, particularly in the interplay between red light and ABA, where critical experimental evidence is lacking.

3.2 Results

3.2.1 Assembly of the Light-Induced Stomatal Opening Signal Transduction Network

The first step in building the model is to construct the regulatory network that represents the system. A network is an abstraction of a system in which each element is represented as a node, and each pairwise interaction or regulatory relationship is represented by an edge. Edges in signal transduction networks are generally directed (meaning that the interaction has a source and a target) and signed (positive or negative). The majority of the known components involved in stomatal opening are proteins, including receptors, enzymes, channels, protein kinases and phosphatases, thus most of the nodes of the network represent proteins. To be able to incorporate the metabolic processes and ion fluxes also involved in stomatal opening, we also include important inorganic compounds, ions, certain biological processes (i.e. photophosphorylation, carbon fixation, stomatal opening) and entities (e.g. mitochondria) as nodes. In some cases, the subcellular localization of a molecule or enzyme can change, making a key difference in the modulation of stomatal opening. In these cases we use multiple nodes, one for each location. Positive edges in the network correspond to activation, up-regulation, or biochemical synthesis, and are represented with a terminating arrowhead, while negative edges indicate deactivation, inhibition, or consumption, and are shown as terminating in a solid circle. The translocation of a protein or the transport of solutes through channels or carriers is also represented by an edge. A relationship stimulated by another component of the network is represented by an edge starting from the stimulus node and incident on the stimulated edge. For instance, malate exits the cytosol and enters the apoplast through active anion efflux channels (AnionCh); this is represented by an edge from

AnionCh incident on the edge starting from cytosolic malate and ending in apoplastic malate. Certain causal regulatory relationships may be mediated by other nodes; a path (a sequence of nodes and edges) is a better representation of such indirect relationships between nodes. I used logical inference to incorporate the components suggested by the totality of relevant experiments to mediate such indirect causal relationships; this process has been formalized previously [19,27]. I distilled more than 85 articles from the literature into 153 edges among 70 nodes, summarized in supporting Table S1 of reference [106]. Supporting Text S2 of [106] provides an illustration of the process of network construction based on the literature, in which the pathway that starts from blue light and ends at the H^+-ATPase is used as an example.

The plant hormone ABA, produced in response to environmental stresses such as drought, opposes the effect of light on guard cells and reduces stomatal apertures [7,10,28,29]. To maintain the focus on stomatal response to light, yet to be able to investigate the cross-talk between different signals, the ABA-response section of the model is a condensed representation of the relevant pathways. I followed two contraction principles to achieve a simpler yet dynamically equivalent representation of the system [30]. Functionally redundant pathways in this section are merged; for instance, the two mechanisms by which NO can elicit calcium release from intracellular stores (CaR) (by cyclic ADP-ribose or by 8-nitro-cyclic guanosine monophosphate [31,32]) are compressed into a single edge from NO to CaR. Further, if the sole known function of an element is to pass on the signal it received, i.e. it has a single incoming activation edge and a single outgoing activation edge, the element is not shown in the model and its upstream regulator is directly connected to its downstream target.

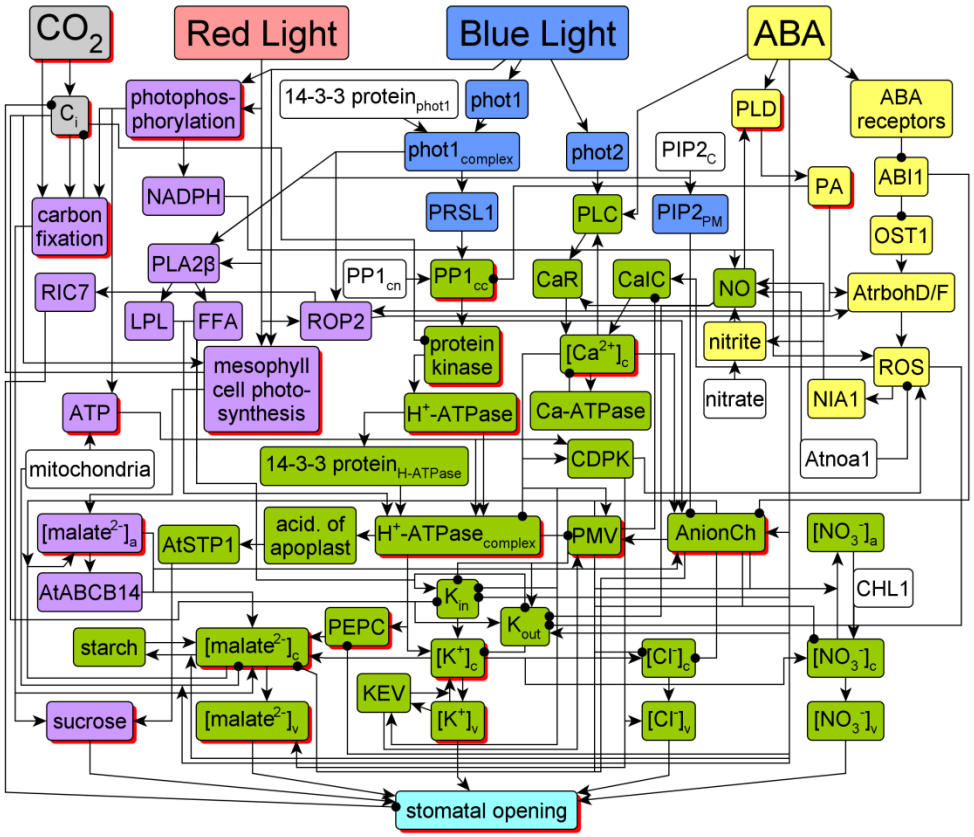


Figure 3-1. Current knowledge of light-induced stomatal opening and its regulation by CO_2 and ABA. The color of the nodes represents their functional connectivity relative to the four signal nodes: CO_2, red light, blue light, and ABA. CO_2 and C_i are colored grey. Nodes that can be activated by blue light alone are colored blue. Nodes that can be activated by either red or blue light are colored purple. Nodes are colored yellow if they respond to the plant hormone ABA, and green if they are affected by both ABA and blue light. Nodes with no upstream effectors (called source nodes) are colored white, stomatal opening is colored teal. I use a red shadow to indicate nodes that are characterized by three or more states in the dynamic model.

To improve the visualization, multiple edges that originate from a single node may start together and bifurcate later toward their individual targets. Similarly, multiple positive edges that end at the same node may merge before reaching the target. Edge bifurcation or merging forms T-shaped junctions, while the crossing of two edges forms plus-shaped junctions.

The full names of the network components denoted by abbreviated node names are: 14-3-3 $\text{protein}_{\text{H-ATPase}}$, 14-3-3 protein that binds to H^+-ATPase; 14-3-3 $\text{protein}_{\text{phot1}}$, 14-3-3 protein that

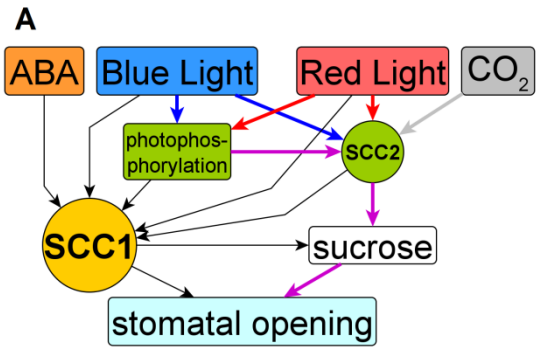
binds to phototropin 1; ABA, abscisic acid; ABI1, 2C-type protein phosphatase; acid. of apoplast, the acidification of the apoplast; AnionCh, anion efflux channels at the plasma membrane; AtABCB14, ABC transporter gene AtABCB14; Atnoa1, protein nitric oxide-associated 1; AtrbohD/F, NADPH oxidase D/F; AtSTP1, H-monosaccharide symporter gene AtSTP1; Ca-ATPase, Ca-ATPases and Ca^{2+}/H^{+} antiporters responsible for Ca^{2+} efflux from the cytosol; CaIC, inward Ca^{2+} permeable channels; CaR, Ca^{2+} release from intracellular stores; carbon fixation, light-independent reactions of photosynthesis; CDPK, Ca^{2+}-dependent protein kinases; CHL1, dual-affinity nitrate transporter gene AtNRT1.1; C_i, intercellular CO_2 concentration; FFA, free fatty acids; H-ATPase, the phosphorylated H-ATPase at the plasma membrane prior to the binding of the H-ATPase 14-3-3 protein; $H\text{-}ATPase_{complex}$, 14-3-3 protein bound H-ATPase; KEV, K^{+} efflux from vacuole to the cytosol; K_{in}, K^{+} inward channels at the plasma membrane; K_{out}, K^{+} outward channels at plasma membrane; LPL, lysophospholipids; NADPH, reduced form of nicotinamide adenine dinucleotide phosphate; NIA1, nitrate reductase; NO, nitric oxide; OST1, protein kinase open stomata 1; PA, phosphatidic acid; PEPC, phospho*enol*pyruvate carboxylase; phot1, phototropin 1; $phot1_{complex}$, 14-3-3 protein bound phototropin 1; phot2, phototropin 2; photophosphorylation, light-dependent reactions of photosynthesis; $PIP2_C$, phosphatidylinositol 4,5-bisphosphate located in the cytosol; $PIP2_{PM}$, phosphatidylinositol 4,5-bisphosphate located at the plasma membrane; PLA2β, phospholipase A2β; PLC, phospholipase C; PLD, phospholipase D; PMV, electric potential difference across the plasma membrane; $PP1_{cn}$, the catalytic subunit of type 1 phosphatase located in the nucleus; $PP1_{cc}$, the catalytic subunit of type 1 phosphatase located in the cytosol; protein kinase, a serine/threonine protein kinase that directly phosphorylates the plasma membrane H-ATPase; PRSL1, type 1 protein phosphatase regulatory subunit 2-like protein1; RIC7, ROP-interactive CRIB motif-containing protein 7; ROP2, small GTPase ROP2; ROS, reactive oxygen species; $[Ca^{2+}]_c$, cytosolic Ca^{2+} concentration; $[Cl^{-}]_{c/v}$, cytosolic/vacuolar Cl^{-} concentration; $[K^{+}]_{c/v}$, cytosolic/vacuolar K^{+} concentration; $[malate^{2-}]_{a/c/v}$, apoplastic/cytosolic/vacuolar $malate^{2-}$ concentration; $[NO_3^{-}]_{a/c/v}$, apoplastic/cytosolic/vacuolar nitrate concentration.

Figure 3-1 represents the resulting network of 70 nodes and 153 edges. The color coding of the nodes signifies the functional connectivity of each node to the four signals, which is based on the existence of paths between a signal and the respective node but is also informed by the specific combinatorial regulation of the node (described in detail in the section "Elements of the dynamic model"). A brief description of the biology represented by the network is as follows; Supporting Text S3 of [106] provides a detailed description of the network. Both red and blue light activate guard cell photophosphorylation, providing adenosine triphosphate (ATP), the primary chemical energy transporter within the cell, for metabolic processes [33]. Subsequent carbon fixation provides sugars, primarily sucrose, as osmotica for guard cell swelling and stomatal opening [34,35]. This pathway is formed by purple colored nodes in the left side of the network. A blue light-specific pathway (blue colored symbols) leads to the activation of the plasma membrane H^+-ATPase [4,36]. H^+-ATPase activity hyperpolarizes the plasma membrane [4], with subsequent uptake of K^+ [37,38] and accumulation of its counterions, Cl^-, NO_3^-, and $malate^{2-}$ [13,33]. These ions also function as osmotica during light-induced stomatal opening [6,39,40]. The stress hormone ABA initiates a signal transduction network (yellow nodes) which ultimately inhibits the plasma membrane H^+-ATPase, inhibits malate synthesis, and induces malate breakdown [2,41-45]. Thus the majority of the nodes in the network (the green-colored nodes) are regulated by blue light and ABA. The twenty-three nodes that have more than two levels in the model are highlighted with a red shadow.

3.2.2 Structural Analysis of the Network

Representing a system with a network reveals important characteristics and interrelationships that have been hidden previously, and enables researchers to test prevailing theories and to identify new hypotheses [46]. I started by looking at the node degree, defined as

the number of edges to which the node is connected, of the 70 nodes. The degree can be broken into the in-degree, i.e. the number of incoming edges (and therefore, of direct upstream regulators), and the out-degree, i.e. the number of outgoing edges (and therefore, of direct downstream targets). The four signal nodes, blue light, red light, CO_2, and ABA, have an in-degree of zero. The node stomatal opening is the only node in the system with an out-degree of zero. Table 1 lists the 10% of nodes with the highest in-degree, out-degree, and total degree, respectively. Most nodes in this list are known key mediators or regulators of light-induced opening. For instance, the node that represents the cytosolic malate^{2-} concentration has the highest in-degree and also the highest total degree in the network. Malate, the major counterion for K^+ in guard cells and a common organic metabolite, is indeed involved in multiple metabolic pathways. The node that represents the 14-3-3 protein-bound H^+ ATPase, H^+-ATPase$_{complex,}$ is also among the nodes with highest in-degree and total degree, indicating its multi-tiered regulation and its important role in determining the membrane potential and hence the flow of multiple ions. The ion channels K_{in}, K_{out}, and anion efflux channels are also among the highly regulated nodes in the system. The stress hormone ABA has the highest out-degree, due to its targeting of multiple nodes in the pathway of blue light-induced stomatal opening and in the ABA signaling network. Cytosolic Ca^{2+} concentration ($[Ca^{2+}]_c$) is an important secondary messenger for both blue light and ABA signaling, as reflected by its high out-degree and total degree. The node PMV, which denotes the potential across the plasma membrane, also has high out-degree and total degree, reflecting its control of channel activities.



B

Edge removed		Number of nodes left in SCC1 (percentage)
starting node	end node	
CaIC	$[Ca^{2+}]_c$	13 (41.9%)
PMV	CaIC	14 (45.2%)
$[Ca^{2+}]_c$	CDPK	18 (58.1%)
CDPK	ROS	18 (58.1%)

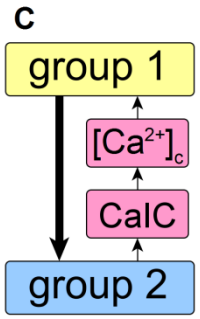


Figure 3-2. Structural analysis of the signaling network. (A) Compressed representation of the network that shows the four signals to the network, two composite nodes that represent SCC1 (which contains 31 nodes) and SCC2 (which contains 3 nodes), as well as photophosphorylation, sucrose, and the output of the network, stomatal opening. The nodes not shown do not affect the network's connectivity and are contracted into the edges shown in black. Five paths do not cross SCC1; they start from blue light, red light, or CO_2, pass through photophosphorylation, merge at SCC2, reach sucrose, and lead to stomatal opening. Signal-specific edges are colored blue (for blue light), red (for red light), grey (for CO_2), the edges shared by both blue and red light are purple. (B) The 4 edges whose removal results in the highest node loss from SCC1. The unperturbed SCC1 contains 31 nodes, which is the basis for the percentage calculation. (C) Sub-structure of SCC1. Group 1 contains 17 nodes, and group 2 contains 12 nodes.

Next, I identified the strongly connected components of the system. A strongly connected component is a group of nodes wherein any node is reachable from any other node through a path (a series of consecutive nodes and edges). Intuitively, a strongly connected component is a closely-knit group of nodes with interwoven feedback that usually forms an important functional module of a network. The stomatal opening network contains three strongly connected components (SCCs), comprising 31 nodes (SCC1), 3 nodes (SCC2) and 2 nodes (SCC3), respectively (Figure 3-2). The 3-node SCC2 represents the interplay amongst C_i and carbon fixation processes in guard cells and mesophyll cells: C_i is required by photosynthesis and photosynthesis lowers C_i in turn. The 2-node SCC3 represents the two directions of transport between apoplastic and cytosolic NO_3^-. The largest SCC signifies the crosstalk between the different signals of the system, since all four signals of the model connect to it. Eight of the thirteen high-degree nodes listed in Table 3-1 are in the largest SCC. Most of the remaining high degree nodes have only outgoing or incoming edges and thus cannot be strongly connected. Twenty-seven nodes, including the nodes of SCC2, can reach the nodes of SCC1 through directed paths. Eleven nodes, including SCC3, can be reached from SCC1 through directed paths. Only a single node, CHL1, is not connected to SCC1 by a directed path.

Degree Types and Values		List of Nodes
In-degree	10	$[malate^{2-}]_c$
	6	H^+-$ATPase_{complex}$, K_{out}, AnionCh, stomatal opening
	5	K_{in}, $[K^+]_c$, $[NO_3^-]_c$
Out-degree	9	ABA
	8	$[Ca^{2+}]_c$, PMV
	6	AnionCh, C_i
	5	phot2
Degree	13	$[malate^{2-}]_c$
	12	PMV, AnionCh
	11	$[Ca^{2+}]_c$
	10	H^+-$ATPase_{complex}$

	9	$[K^+]_c$, ABA

Table 3-1. The top 10% of nodes in terms of in-degree, out-degree, and total degree in the network. For example, the node AnionCh has an in-degree of 6, an out degree of 6, and a total degree of 12.

There are 19,436 simple paths (i.e., paths with no repeated nodes) between the four signal nodes and stomatal opening. The vast majority of these paths pass through SCC1. Remarkably, five important paths bypass SCC1 (see Figure 3-2A). The five paths start from blue light, red light and CO_2, respectively, pass through photophosphorylation and SCC2, and then through sucrose, and finish at stomatal opening. Photophosphorylation and SCC2 together represent the photosynthetic carbon fixation processes in guard cells and mesophyll cells. Two conclusions can be drawn based on the existence of these SCC1-bypassing paths: i) The five paths represent photosynthetic carbon reduction pathways. ABA does not inhibit photosynthesis in mesophyll cells [47], and there is no indication that ABA would inhibit guard cell photosynthesis. Thus there is no current wet bench evidence that ABA would be able to affect the accumulation of sucrose via guard cell photosynthesis (see Figure 3-2A). ii) Based on current knowledge, sucrose accumulation does not introduce any feedback into the system.

I next ranked the 60 edges of SCC1 according to their importance to SCC1 integrity. If a strongly connected component is densely connected, the loss of a single edge should not affect the reachability of node pairs in it. However, if after the loss of an edge certain nodes can no longer reach other nodes, or cannot be reached from other nodes, they are no longer part of the strongly connected component and thus the number of nodes in the strongly connected component decreases. Supporting Table S2 of [106] provides information on the effects of removal of each edge. Removal of any one of 26 edges led to no change in the composition of SCC1. Loss of any one of 19 edges led to minimal changes, i.e. the loss of a single node from SCC1. Among the edge

removals that do induce significant breakdown, the four listed in Figure 3-2B lead to loss of more than 40% of the nodes in SCC1. All four edges are closely related to $[Ca^{2+}]_c$, indicating the critical role of $[Ca^{2+}]_c$ in the formation of SCC1. A closer examination revealed that SCC1 contains two smaller groups of strongly connected nodes (Figure 3-2C). Group 1 contains 12 nodes from the ABA signaling pathways. Group 2 contains 17 nodes, including the H^+-ATPase, its four regulators, nodes denoting major ions, and PMV, which are major mediators of blue light signaling. Seven edges connect group 1 to group 2, which is the reason why the nodes in group 2 are colored green in Figure 3-1. However, there is a single path from group 2 back to group 1, mediated by CaIC and $[Ca^{2+}]_c$. The loss of any of the nodes or edges involved in this path results in a major breakdown of SCC1 (Figure 3-2B). The fact that both groups are strongly connected with $[Ca^{2+}]_c$ indicates that $[Ca^{2+}]_c$ bridges the signaling between blue light and ABA. This conclusion is corroborated experimentally [48-53]. Indeed, it is known that $[Ca^{2+}]_c$ is an important secondary messenger in both blue light [53,54] and ABA signaling [55,56]. The strongly connected component analysis offers additional insight into the role of $[Ca^{2+}]_c$ and reveals that it is a key participant in a feedback loop formed by these pathways.

3.2.3 Elements of the dynamic model

The signal transduction network described in Figure 3-1 forms the basis of the dynamic model of light-induced stomatal opening. The dynamic model characterizes each node with a state variable (which we will also refer to as "level") and with a regulatory function (also called transfer function) that indicates the future state of the node as a function of the current state of its regulators. Iterative determination of each node's state from a suitable initial condition yields the dynamic behavior of the whole system. Importantly, the global dynamics of the whole system is an emergent

property that cannot be directly controlled by the modeler but arises from the local dynamics (the regulation of each component).

I developed a discrete dynamic model in which the nodes are assigned two or more qualitative levels. I aimed to employ the minimal number of levels that was sufficient to describe the experimentally observed relative outcomes for various conditions (e.g. combinations of signals and manipulations of node states). The two possible levels of binary nodes (0 and 1) can be interpreted as "OFF", "low" or "inactive", versus "ON", "high", or "active". Three levels can be interpreted as "low", "medium" and "high". The benefit of having three-level nodes is most evident when three qualitatively different categories of values are observed under three or more different experimental conditions, e.g. stomatal opening under red light alone, blue light alone or under dual beam. For such scenarios, having nodes with only two levels would force the grouping of qualitatively different values, and therefore lead to information loss. Among the observations that necessitated the use of more than two levels are the synergistic (stronger than additive) effect between red light and blue light in malate formation [13] and stomatal opening [3,5,9,39,57-59], and the complex behavior of $[Ca^{2+}]_c$ as a secondary messenger during blue light-induced stomatal opening [50,51,60] and ABA-induced stomatal closure [43,48,49,61]. In addition, the osmotic potential difference across the plasma membrane that leads to stomatal movement results from the totality of all solutes, whose effect is biophysically additive. In the model, 47 nodes have two levels, nine nodes have three levels (including photophosphorylation, carbon fixation, $[Ca^{2+}]_c$, CO_2), two nodes have four levels (ATP, C_i), three nodes have five levels (Protein Kinase, H^+-$ATPase_{complex}$, PMV) and nine nodes have more than five levels (including protein kinase, H^+-$ATPase_{complex}$, $[K^+]_c$, $[malate^{2-}]_c$, stomatal opening). Stomatal opening, in particular, has more than 20 reachable levels, ranging from 0 to 14.01. The numerical values of these levels are not

meaningful in isolation; rather, their relationships are reflective of the experimentally observed relative outcomes. Supporting Text S1 of [106] provides a listing of all node levels.

The four signals of the model were assigned a set of levels that represent a particular experimental condition (light condition, CO_2 concentration and ABA presence or absence). The possible levels of blue light, red light and ABA are ON (1) or OFF (0), indicating their presence or absence. CO_2 has three levels, 0 (reduced CO_2), 1 (ambient atmospheric CO_2), and 2 (high CO_2). The signal levels can be externally changed, e.g. to simulate a light pulse experiment. The 64 internal nodes were chosen to have an initial state of 1 (7 nodes) or 0 (57 nodes) based on experimental information. Supporting Text S1 of [106] describes these initial states and their justification.

Time is discretized into steps in the model. In one time step, the state of each node is updated according to the transfer function assigned to it [19]. I followed random order asynchronous update [62]. A random permutation of the nodes (except the node stomatal opening) is first established at the beginning of each time step, and then all nodes are updated according to this sequence. Stomatal opening, as the sole output of the model, is always updated last within each time step. This algorithm effectively implements a random sampling of process durations. We have chosen this random sampling due to the scarcity of experimental data on relative reaction speeds of signaling pathways and on the timing of specific intracellular events. The degree of randomness can be reduced as timing information becomes available. A delay of 10 time steps is implemented for the node sucrose (see the transfer function for the node sucrose in Supporting Text S1 of [106]). I determined empirically that for the network a total of 18 time steps in each simulation is sufficient for all components to reach a time-invariant state (steady state or, for a minority of nodes such as $[Ca^{2+}]_c$, sustained oscillation).

The transfer function of a target node indicates the future state of the target node as a function of the current states of the nodes that have a directed edge impinging on the target. The transfer functions were developed with information from the literature, such as the state of the target node when one of its regulators is knocked out, and basic biochemical or physical principles when applicable. I aimed to construct the simplest transfer functions to minimize the number of unknown parameters in the model. The transfer functions combine logic clauses (using the Boolean operators NOT, AND, OR) with addition, subtraction, and multiplication. This approach enables a more detailed and accurate representation than traditional Boolean models, while maintaining simplicity and using few parameters. Two examples of transfer functions are given in the Materials and Methods section, and Supporting Text S1 of [106] provides a full list of the transfer functions and their justifications.

I tested different numbers of replicate simulations, and found that 2,000 replicate simulations were sufficient for a high reproducibility of the results (see Supporting Text S4 of [106]). I also demonstrated that the model is robust against uncertainty in the update rules without losing its sensitivity to new information on critical nodes (see Supporting Text S4 of [106]). For each experimental condition studied, a total of 2,000 simulations was performed, and for each node, the activity level averaged over all simulations is reported. Experimental condition refers to the level of the four signals and/or any other elements of the system that might be silenced to represent knockout (KO) experiments or made constitutively active; these factors are then invariant across all 2,000 runs.

Importantly, since the input to the model is local (the relationships among pairs of nodes, see Supporting Table S1 of [106]), an agreement between the global dynamic results of simulations from the model and wet bench results is not an inherent property of the model. As shown below,

however, the model does in fact successfully reproduce known dynamic features exhibited by stomatal opening under various conditions, providing strong support for the validity of the model.

3.2.4 The model recapitulates and elucidates wild type responses to light

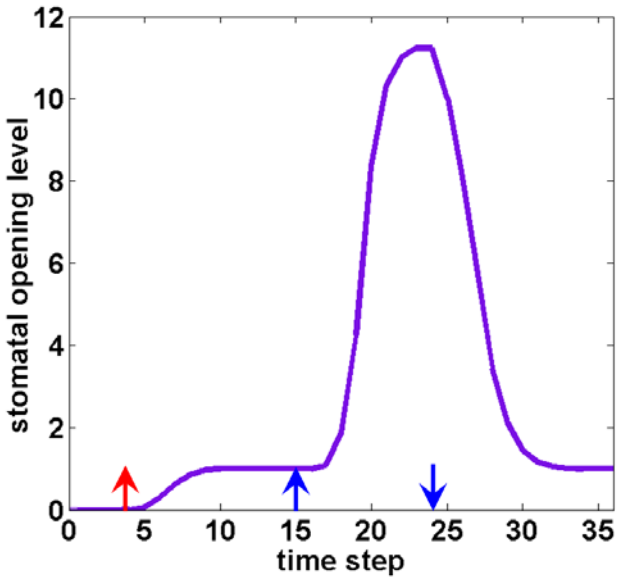
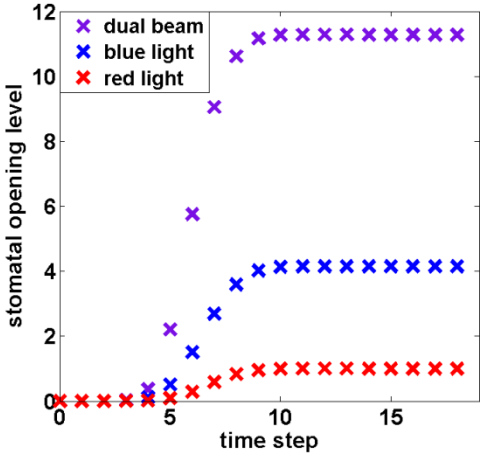


Figure 3-3. Simulation of stomatal opening level in response to a sequence of light conditions. The arrows with corresponding colors and directions signify the imposition (upward) or removal (downward) of a specific light signal. The system is in darkness at time step 0; a red light signal is added at step 4; a blue light signal is turned on at step 15 and off at step 24. The blue light signal induces a sharp increase in the stomatal opening level. The stomatal opening level gradually returns to the red light-induced steady state level after the blue light pulse.

I started by comparing the model's results to experiments under different qualities of light in ambient air. In signature experiments that investigated the roles of red and blue light in stimulating stomatal opening, leaves were illuminated with constant background red light upon which a short blue light pulse was superimposed. The stomatal conductance increased slightly in response to the red light, then displayed dramatic transient increase in response to the blue light

pulse [5,9,57,58]. As depicted in Figure 3-3, the model successfully reproduces this temporal pattern of stomatal opening.

A



B

light quality	maximum stomatal opening	$t_{50\%}$	$t_{95\%}$	maximum malate levels	maximum H^+-$ATPase_{complex}$ activation level	composition of major osmotica	
						ions (K^+, Cl^-, NO_3^-, $malate^{2-}$)	sucrose
dual beam	11.28	6	10	6.5	6	82.5%	17.5%
blue light	4.15	7	10	1.5	2	75.9%	24.1%
red light	1	7	9	0	0	0.0%	100.0%

Figure 3-4. Simulation of stomatal opening under different conditions of light quality in ambient air. (A) Mean stomatal opening levels as a function of time step from 2,000 simulations. Purple: dual beam (blue light = red light = 1, CO_2 = 1); blue: blue light (blue light = 1, red light = 0, CO_2 = 1); red: red light (blue light = 0, red light = 1, CO_2 = 1). The standard error of the mean for the stomatal opening level is smaller than the symbols, and is consequently not shown. (B) Summary table of results for several simulated variables. The first three columns summarize the results shown in (A) indicating the maximum (steady-state) opening level, the number of time steps at which 50% of simulations reach 50% of the maximum level ($t_{50\%}$) and the number of time steps at which 95% of simulations reach 95% of the maximum level ($t_{95\%}$). The next two columns indicate the maximum malate levels and the maximum activation levels of the H^+-$ATPase_{complex}$. The two right-most columns present the contribution of different osmotica (ions vs. sucrose) to stomatal opening in response to different light qualities.

I simulated wild type stomatal responses to sustained light in ambient air, as illustrated in Figure 3-4A. The specific combination of signals for each curve (blue light, red light, or dual beam) is initiated at time step 0 and maintained throughout the simulation. All three time courses of average stomatal opening levels (over the 2000 simulations) have similar sigmoidal shapes. I consistently observed sigmoidal timecourses for stomatal opening and other variables and in the following summarize them by three parameters (Figure 3-4B): the maximal (steady state) value of the mean level, the number of time steps at which 50% of simulations reach 50% of the maximal (steady state) level ($t_{50\%}$), and the number of time steps at which 95% of all simulations reach 95% of the maximal level ($t_{95\%}$). In the presence of both blue and red light, the average stomatal opening level reaches a maximum of ~11.28 in ~10 steps, whereas red light alone only generates an opening level of 1.00. Notably, blue light, with an opening level of 4.15, is more effective than red light in inducing opening, which is consistent with experimental observations of stomatal apertures [3,5]. A synergistic action of red and blue light on stomatal opening, which has been observed experimentally [5,9,39,57], is reproduced in Figure 3-4A: the stomatal opening level under both blue and red light (dual beam) is larger than the sum of opening levels under each type of light alone.

Malate, a common organic compound found in plants, is one of the major counterions for K^+, causing guard cell swelling and stomatal opening. The action spectrum of malate formation shows a synergistic action between red and blue light, i.e. the malate synthesis level under blue light with a red light background (dual beam) is higher than the sum of levels under each type of light (red or blue) alone [13]. This provides a valuable criterion to evaluate the model. Simulation results presented in Figure 3-4B clearly indicate that the maximal malate level under dual beam illumination is higher than the sum of maximal levels accumulated under individual light qualities.

The result that malate has no observable accumulation under red light alone is also in accordance with experiments [35].

Also listed in Figure 3-4B is the maximum activation level of the H^+-$ATPase_{complex}$ obtained in the model under each light condition with ambient air. The proton pump, H^+-ATPase, is responsible for the plasma membrane polarization status and for concurrent ion flows. The model indicates that the H^+-$ATPase_{complex}$ is activated to the highest degree under a dual beam, to a significant degree under blue light alone, and is inactive under red light alone. Experimental evidence on the activation of the H^+-ATPase under red light alone is mixed (see section 2.3 Discussion). The model supports the conclusion that in ambient air the H^+-ATPase is not significantly activated under red light. The result that blue light alone can activate the proton pump is consistent with multiple experiments [4,36,63-65]. The model predicts synergy between red and blue light in the activation of the proton pump, and it suggests that this synergy is one of the mechanisms that underlies the synergy between red and blue light in stomatal opening and malate accumulation.

Figure 3-4B also presents the model's prediction of the relative contribution of the two major types of osmotica, ions (K^+ and its counterions) and sucrose, to the osmotic potential under different light qualities in ambient air. These relative contributions are normalized such that their sum is 100%; see Supporting Text S1 of [106] for the detailed definition of the contribution of each osmoticum to osmotic potential in the model. The model indicates that ions are the predominant osmoticum being accumulated in response to dual beam or blue light (82.5% and 75.9%, respectively), whereas sucrose is the sole osmoticum responsible for red light-induced stomatal opening. These results agree with experimental findings: ion accumulation was observed to take place predominantly under white light or blue light, and is nearly non-observable under red

light under ambient CO_2 conditions [6,35,39]; sucrose accumulation takes place under either blue or red light [35].

3.2.5 The model explains the effect of external CO_2 levels

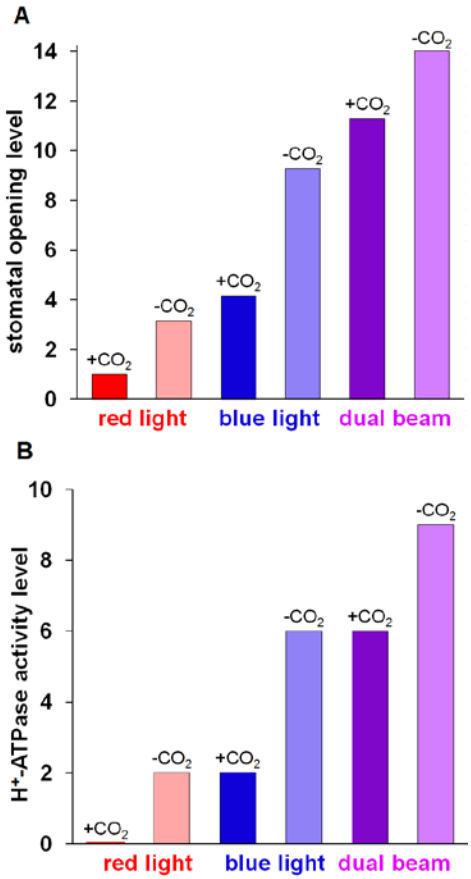


Figure 3-5. The effect of CO_2-free air on light-induced stomatal opening and H^+-ATPase activity. Simulations of (A) maximum stomatal opening level, and (B) maximum H^+-ATPase activity level in air with moderate CO_2 concentration ($+CO_2$) compared to CO_2-free air ($-CO_2$) under different light conditions. Red color indicates red light, blue color indicates blue light, purple color indicates dual beam. Darker colors represent air with moderate CO_2, and lighter colors represent CO_2-free air. (A) Stomatal opening is significantly enhanced by CO_2-free air under all light conditions. (B) The H^+-ATPase activity pattern parallels that of stomatal opening levels in having higher levels in the absence of CO_2.

I investigated the effect that different levels of CO_2, another input signal to the model, has on stomatal opening induced by different qualities of light. The CO_2 content in the ambient atmosphere affects light-induced stomatal opening. Air with lower CO_2 concentration or CO_2-free air was shown to promote white light-induced stomatal opening [1], blue light-induced stomatal opening [5,58], and red light-induced stomatal opening [40]. The model captures the enhancement of stomatal opening levels by low CO_2 under all light conditions (Figure 3-5A). The simulations also indicate that the pattern of the maximal H^+-ATPase activity in response to different light and CO_2 conditions parallels that of stomatal opening (Figure 3-5B). The model thus predicts that the H^+-ATPase activity level is promoted by CO_2-free air compared to ambient air under all light conditions, and suggests that the promotion of H^+-ATPase activity level may contribute to the enhancement of stomatal opening levels by CO_2-free air.

3.2.6 The model recapitulates perturbation scenarios

In order to further test the validity of the model, I next investigated a number of perturbation scenarios.

A

Treatment	Maximum Stomatal Opening	$t_{50\%}$	$t_{95\%}$
Dual Beam	11.28	6	10
Dual Beam+DCMU	3.15	6	9
Blue Light	4.15	7	10
Blue Light+DCMU	1.58	7	10
Red Light	1	7	9
Red Light+DCMU	0	0	0

B

Treatment	Maximum Levels		$t_{50\%}$	$t_{95\%}$
Dark	Stomatal Opening	0	0	0

	K^+ Uptake	0	0	0
Fusicoccin+Dark	Stomatal Opening	9	3	4
	K^+ Uptake	9	3	4

Table 3-2. Simulated effects of DCMU and fusicoccin. (A) The effect of DCMU on stomatal opening under different light conditions. (B) Stomatal opening and K^+ uptake induced by fusicoccin in darkness.

DCMU, a photosynthetic inhibitor, completely inhibits red light-induced stomatal opening, but only partially inhibits blue light-induced stomatal opening [3]. The simulation of the DCMU effect (via maintaining the node photophosphorylation at level 0) is consistent with these experimental observations: the stomatal opening level drops from 1 to 0 under red light, indicating a total inhibition, while the same disruption has a partial effect on stomatal opening induced by a dual beam or by blue light (see Table 3-2).

Fusicoccin is a fungal toxin that stimulates K^+ uptake in guard cells and causes stomatal opening in darkness [66,67]. Fusicoccin has been widely used as a physiological tool to investigate guard cell signaling [68]. Fusicoccin activates the plasma membrane H^+-ATPase via a mechanism that involves inactivation of an autoinhibitory domain [69,70]. I simulated the effect of fusicoccin on the H^+-ATPase by fixing the state of the H^+-ATPase at its maximum activation level. The simulation indicated that without fusicoccin, stomatal opening and K^+ levels remain 0 in darkness (see Table 3-2). This is due to the absence of H^+-ATPase activation in the dark. When fusicoccin is present, stomata open despite the absence of light, and K^+ increases in the darkness as well. The simulation suggests that the H^+-ATPase, when activated by fusicoccin in the dark, leads to the hyperpolarization of the plasma membrane and the subsequent activation of K^+ uptake channels. The accumulation of K^+ and its counterions in the guard cell is the cause for stomatal opening in

the dark in the presence of fusicoccin. Both behaviors are consistent with experimental findings [66,67].

3.2.7 The model reveals the relative contributions of different osmotica

K^+ and its counterions, and sugars, mostly sucrose, are the two types of primary osmotica contributing to stomatal opening. As presented earlier, light quality is one of the determining factors of osmotic composition (Figure 3-4B). In addition, varying the environmental CO_2 concentration was shown to have an effect on osmotic composition as well: CO_2-free air or air with low CO_2 concentration was observed to induce stomatal opening accompanied by K^+ uptake in response to red light [40]. I systematically investigated the effect of different combinations of light qualities and CO_2 concentrations on the contribution of each type of osmoticum during stomatal opening (Table 3-3). The results indicated that in ambient air, ion accumulation is the predominant mechanism leading to stomatal opening under dual beam or blue light, while sucrose is the major osmoticum during red light-induced stomatal opening. In the case of CO_2-free air or air with reduced CO_2 concentration, the model corroborates Olsen et al. (2002) on the importance of K^+ uptake during red light-induced stomatal opening [40], and predicts the absence of sucrose and the predominance of ion accumulation as an osmoticum under all light qualities. The model predicts that ion accumulation is severely suppressed in air with elevated CO_2 concentration, and that under high CO_2 concentration, sucrose is the primary osmoticum responsible for stomatal opening under all light conditions.

CO_2 Content	Light	Phenotype	Maximum Stomatal Opening	$t_{50\%}$	$t_{95\%}$	Composition of Osmotica	
						Ions (K^+, Cl^-, NO_3^-, $malate^{2-}$)	Sucrose

Ambient CO_2	Dual Beam	Wild Type	11.28	6	10	82.5%	17.5%
		H^+-ATPase KO	2	7	10	0.0%	100.0%
		Sucrose Depletion	9.28	6	9	100.0%	0.0%
	Blue Light	Wild Type	4.15	7	10	75.9%	24.1%
		H^+-ATPase KO	1	7	10	0.0%	100.0%
		Sucrose Depletion	3.15	7	10	100.0%	0.0%
	Red Light	Wild Type	1	7	9	0.0%	100.0%
		H^+-ATPase KO	1	7	9	0.0%	100.0%
		Sucrose Depletion	0	0	0	—	—
CO_2-Free Air	Dual Beam	Wild Type	14.01	6	8	100.0%	0.0%
		H^+-ATPase KO	0	0	0	—	—
		Sucrose Depletion	14.01	6	8	100.0%	0.0%
	Blue Light	Wild Type	9.28	6	8	100.0%	0.0%
		H^+-ATPase KO	0	0	0	—	—
		Sucrose Depletion	9.28	6	8	100.0%	0.0%
	Red Light	Wild Type	2	6	7	100.0%	0.0%
		H^+-ATPase KO	0	0	0	—	—
		Sucrose Depletion	2	6	7	100.0%	0.0%
Elevated CO_2	Dual Beam	Wild Type	2	7	9	0.0%	100.0%
		H^+-ATPase KO	2	7	9	0.0%	100.0%
		Sucrose Depletion	0	0	0	—	—
	Blue Light	Wild Type	1	7	9	0.0%	100.0%
		H^+-ATPase KO	1	7	9	0.0%	100.0%
		Sucrose Depletion	0	0	0	—	—
	Red Light	Wild Type	1	7	9	0.0%	100.0%
		H^+-ATPase KO	1	7	9	0.0%	100.0%
		Sucrose Depletion	0	0	0	—	—

Table 3-3. Simulated stomatal opening levels and osmotic compositions under various conditions of light, CO_2, and node disruptions. The CO_2 conditions studied are: ambient CO_2 concentration (CO_2 = 1, top of the table), CO_2-free air (CO_2 = 0, middle), and elevated CO_2 concentration (CO_2 = 2, bottom). Simulated H^+-ATPase knockout (H^+-$ATPase_{complex}$ = 0) severely impairs stomatal

opening in the cases where ions are the predominant osmotica, e.g. under CO_2-free air. Computationally imposed sucrose depletion (sucrose = 0), on the other hand, inhibits stomatal opening in cases where sucrose is the major osmoticum, e.g. under elevated CO_2 concentration.

I further probed the importance of different types of osmotica by computationally imposing an inhibition of sucrose accumulation (sucrose = 0) or by virtually knocking out the H^+-ATPase (H^+-$ATPase_{complex}$ = 0). The model indicated that in ambient air (Table 3-3, top), the H^+-ATPase plays a more important role in dual beam- or blue light-induced stomatal opening than in red light-induced stomatal opening. Conversely, I found that sucrose is more important for red light-induced stomatal opening than for stomatal opening under dual beam or blue light, since its knockout completely inhibits red light-induced stomatal opening while the inhibition is partial for dual beam and blue light. Under CO_2-free air (Table 3-3, middle), H^+-ATPase activity is more critical than sucrose accumulation for stomatal opening under all light conditions. In fact, keeping sucrose at value 0 computationally has no effect on stomatal opening in CO_2-free air. In air with elevated CO_2 (Table 3-3, bottom), the proton pump and henceforth ion accumulation are suppressed, making sucrose the predominant osmoticum for stomatal opening under all light conditions. These results also confirm that the activity of the proton pump is the primary driving force for ion accumulation during stomatal opening [2,71,72].

Light	Phenotype	Maximum Stomatal Opening	$t_{50\%}$	$t_{95\%}$
Dual Beam	Wild Type	11.28	6	10
	K_{in} KO	2	7	9
Blue Light	Wild Type	4.15	7	10
	K_{in} KO	1	7	9
Red Light	Wild Type	1	7	9
	K_{in} KO	1	7	9

Table 3-4. The effect of inward K^+ channel knockout on stomatal opening under different light conditions predicted by the model. K_{in} knockout has a larger effect in dual beam- and blue light-induced stomatal opening, while it has no observable effect on red light-induced stomatal opening.

Consistent with the difference in the types of osmotica mediating blue light or red light-induced stomatal opening [6,35], a K_{in} knockout displayed a more severe reduction of opening level in white light and blue light than in red light [73]. This phenomenon is also captured by the model (Table 3-4): the simulated white or blue light-induced stomatal opening level decreases dramatically when K_{in} is forced to be 0, but this disruption does not affect the red light-induced stomatal opening level.

3.2.8 The model predicts the effects of single knockouts

I performed a systematic compilation and comparison of available experimental observations with results generated by the model in a simulation of the experimental conditions. These conditions included different light and/or CO_2 and/or ABA stimuli and the manipulation of node states by genetic modifications or pharmacological interventions. Sixty-six comparisons were made in total (see Supporting Table S4 of [106]), out of which 64 instances exhibited qualitative consistence between experimental observations and simulation results—a successful validation rate of 97%.

The model's consistency with known experimental evidence enables confident prediction of new phenotypes. It takes a significant amount of time and effort for experimentalists to investigate the effect of the genetic knockout of even a single element *in vivo*. In contrast, a compilation of the phenotypes of all the single-node knockout phenotypes can be readily obtained *in silico*, and can then be used to inform and prioritize experiments.

Opening Level As a Percentage of WT Opening		0-5%	5%-15%	15%-25%	25%-35%	35%-45%	45%-55%	55%-65%	65%-75%	75%-85%	85%-95%	95%-100%	100%-105%
Light Quality	Air Condition	Percentage of Single Knockouts in Each Bin											
Dual Beam	Moderate CO_2			17.2%	1.6%	1.6%			1.6%	9.4%	1.6%	64.1%	3.1%
	CO_2-free	17.2%			3.1%				7.8%	1.6%	1.6%	65.6%	3.1%
	High CO_2	4.7%										95.3%	
Blue Light	Moderate CO_2			18.8%		1.6%		1.6%		6.3%	3.1%	68.8%	
	CO_2-free	17.2%			1.6%		3.1%		6.3%	1.6%	1.6%	65.6%	3.1%
	High CO_2	4.7%										95.3%	
Red Light	Moderate CO_2	4.7%										95.3%	
	CO_2-free	17.2%					3.1%		6.3%	1.6%	1.6%	70.3%	
	High CO_2	4.7%										95.3%	

Table 3-5. The distribution of predicted light-induced stomatal opening levels for single node knockouts. Each simulated knockout mutant's opening level is expressed as a percentage of the wild type opening level for the corresponding light quality and CO_2 condition. ABA is absent in all simulations. The opening levels are binned into 12 ranges, indicated in the header of the table. Each entry indicates the percentage of the 64 knockouts in each opening category. The entry is left blank if no knockout mutant opening level falls in the corresponding range. In the moderate CO_2 cases, an average of 77.1% of all single knockouts maintains an opening level that is less than 5% different from wild type opening, with less than 2% displaying major inhibition ($\geq$ 95%) of opening, demonstrating the robustness of the system against single node losses. Single node-knockouts have a larger impact on stomatal opening under CO_2-free air: an average of 69.3% of all single node-knockouts maintains an opening level less than 5% different from wild type opening, while 17.2% of all knockouts result in major inhibition of stomatal opening in all light conditions. Under high CO_2 condition, interestingly, all light conditions exhibit identical knockout opening pattern: 95.3% of all single node knockouts display close to wild type opening, and 4.7% display major inhibition.

In vivo, a null phenotype is realized by creating knockout mutants, or by introducing a pharmacological suppressor of a certain element. *In silico*, this is achieved by keeping the level of

the ‘knocked-out’ node at 0. I systematically investigated the effect of the knockout of a single node from the system in the following three light conditions: dual beam (blue light = red light = 1), blue light alone (blue light = 1, red light = 0), and red light alone (blue light = 0, red light = 1), and three atmospheric conditions: normal air with moderate CO_2 concentration (CO_2 = 1), CO_2-free air (CO_2 = 0), and high CO_2 air (CO_2 = 2). ABA was set as absent (ABA = 0) in these simulated knockouts, in which each of the 64 internal nodes was individually eliminated *in silico*. Table 3-5 lists the distribution of stomatal opening levels of the knockout phenotypes as a percentage of the wild type opening, which was equated to 100%. In all three ambient air cases, the majority of the knockouts (67.2% for dual beam, 68.8% for blue light, and 95.3% for red light) maintained an opening level within 5% deviation from wild type opening, demonstrating the robustness of the system against single node loss.

The knockout of single nodes has a larger impact on stomatal opening under CO_2-free air, predominantly due to the inactivity of photosynthetic carbon fixation pathways in this condition, making H^+-ATPase activation and the accumulation of ions crucial to stomatal opening. A smaller fraction of cases maintained an opening within 5% of wild type opening (68.7% for dual beam, 68.7% for blue light, and 70.3% for red light), and 17.2% of all phenotypes resulted in an inhibition of 95% or more of wild type stomatal opening level in all light conditions under CO_2-free air. Under the high CO_2 condition, since the proton pump H^+-ATPase activity is greatly suppressed, stomatal opening under all three light conditions is solely dependent on photosynthesis and sucrose accumulation. Therefore, stomatal opening of knockout phenotypes under the three light conditions have an identical pattern: 95.3% stay close to wild type opening level, and 4.7% display total inhibition of opening.

Interestingly, knocking out the small G protein ROP2 or RIC7 induced a stomatal opening level higher than wild type opening under three different light and CO_2 conditions (see Table 3-5, 100%-105%). This model result recapitulates the experimental observation that ROP2 and recruited RIC7 inhibit stomatal opening in wild-type plants, thus providing a protection mechanism against excessive opening [74]. Supporting Table S3 of [106] provides a full list of stomatal opening levels for each simulated node knockout in each of the nine conditions.

3.2.9 The model captures the known effects of ABA on stomatal opening and identifies a knowledge gap

	ABA Absent			ABA Present		
Light Treatment	Maximum Stomatal Opening	$t_{50\%}$	$t_{95\%}$	Maximum Stomatal Opening	$t_{50\%}$	$t_{95\%}$
Dual Beam	11.28	6	10	2	7	9
Blue Light	4.15	7	10	1	7	9
Red Light	1	7	9	1	7	9

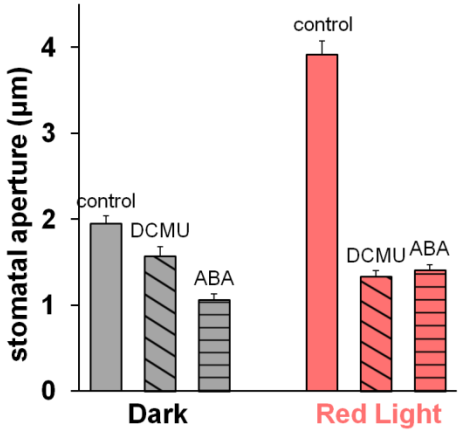
Table 3-6. Predicted stomatal opening level under different qualities of light in the absence or presence of ABA. The presence of ABA leads to a dramatic decrease in the maximal stomatal opening level under dual beam or blue light but ABA has no effect on stomatal opening under red light.

It is known that under simultaneous presence of white light and ABA, the latter functions through several secondary messengers, e.g. $[Ca^{2+}]_c$ [75,76] and pH_c [77], to inhibit light-induced stomatal opening. The model reproduced this effect as shown in Table 3-6. ABA decreased stomatal opening under combined blue and red light from 11.28 to 2, and stomatal opening under blue light decreased from 4.15 to 1. Unexpectedly, however, the model predicted that ABA had no inhibitory effect on red light-induced opening, which remained at level 1 regardless of ABA.

In the course of the construction of the network of guard cell secondary messengers of light and ABA signaling, I found 30 components (nodes) wherein regulation of the node by both blue light and ABA had been reported or could be inferred, consistent with experimental evidence that ABA inhibits blue light-stimulated stomatal opening [17], and the model clearly indicated ABA-inhibition of blue light stimulated stomatal opening (Table 3-6, blue light). By contrast, we found no nodes for which regulation of the node by both red light and ABA had been reported. Accordingly, in the model, ABA is predicted to have no effect on red light-induced stomatal opening (Table 3-6, red light). This prediction led me to extensively peruse the literature for experiments in which ABA inhibition of red light-induced stomatal opening in isolated epidermes had been explicitly assessed, but no such reports were found. This absence of studies perhaps reflects the general (but untested) belief that ABA is able to inhibit light-induced stomatal opening regardless of the wavelength of light. The identification of this unaddressed question exemplifies how codification of extant knowledge into network models can suggest key new experiments.

3.2.10 Experimental test of the effect of ABA on red light-induced stomatal opening

A



B

light condition		dark			red light		
treatment		control	DCMU	ABA	control	DCMU	ABA
maximum stomatal opening level	original model	0	0	0	1	0	1
	modified model	0	0	0	1	0	0

Figure 3-6. The effect of ABA and DCMU on red light-induced stomatal opening. (A) Experimental measurement of stomatal apertures in isolated epidermal peels of *Vicia faba* under different conditions. Qualitatively, the apertures can be categorized into two opening levels: red light yields a high opening level, and opening levels under all other conditions can be considered as low. (B) Simulated stomatal opening levels. In the original model, stomatal opening levels under red light and red light + ABA treatments are high (level 1); all other treatments yield low (level 0) opening levels. If the model is modified with an inhibition by ABA of sucrose accumulation, red light is the only condition that yields a high opening level; all other conditions have low opening levels. This is in qualitative agreement with the experimental results shown in (A).

Accordingly, our experimental collaborator Xiaofen Jin in Prof. Assmann's laboratory assessed the effect of ABA on red light-induced stomatal opening in *Vicia faba* epidermal peels (see Materials and Methods). Significant inhibition by ABA of red light-induced stomatal opening was found (Figure 3-6A). The inhibition of stomatal opening by the photosynthetic inhibitor, DCMU, was also observed, consistent with a previous report [3].

Qualitatively, the average stomatal aperture values from the wet bench experiments can be divided into two groups: those that have a high value, of which red light treatment is the only instance, and those that have a low value, which contains all the other experimental conditions. Figure 3-6B shows the stomatal opening levels obtained from the model for the same conditions as those studied experimentally. As the model, constructed based on available knowledge, lacks a mechanism through which ABA can inhibit red light-induced opening, combined action of red light and ABA results in a high opening level equal to that of red light alone, while all other cases have low opening levels. The model predictions are qualitatively consistent with the experimental findings with the exception of the case of combined red light and ABA input. The discrepancy between the model and experimental results in the latter case points out a missing piece of the current knowledge base of the model: a mechanism through which ABA can inhibit red light-induced stomatal opening.

3.2.11 A hypothesis on ABA inhibition of red light-induced stomatal opening

The question 'how does ABA inhibit red light-induced stomatal opening?' remains open. Since sucrose is the major osmoticum accumulated under red light in ambient CO_2, a natural first hypothesis is that ABA inhibits sucrose accumulation. This hypothesis is supported by the result that there are only three nodes whose individual knockout abolishes red light-induced stomatal opening: light reaction, carbon fixation and sucrose (see Supporting Table S3 of [106]). These nodes form a linear path from ABA to sucrose (see Figure 3-2). ABA could disrupt the reaction cascade through which sucrose is generated, or cause the conversion of sucrose into starch, or promote sucrose catabolism within the guard cell or its efflux from the cell.

To explore the explanatory power of a putative ABA inhibition of sucrose accumulation, I modified the model by adding an inhibitory edge from ABA to sucrose and adding the Boolean clause "*And Not* ABA" to the existing rule for sucrose. The simulation result of this modified model is shown in Figure 3-6B. Notably, after this modification, ABA is able to inhibit red light-induced stomatal opening. The qualitative response pattern across all treatments matches the experimental results. Further, I assessed the impact of the putative inhibitory edge from ABA to sucrose on the overall performance of the model by comparing all the simulation results obtained from the modified model with those obtained from the original model. I was able to confirm that all results stay either identical (e.g., for conditions where the ABA signal is absent) or qualitatively consistent (for conditions where the ABA signal is present, see Supporting Table S5 of [106]). Importantly, this exemplifies how discrete models, even in the absence of complete knowledge of all interactions and of temporal signaling dynamics, can be readily employed to test new hypotheses and putative pairwise relationships between components of the system.

3.3 Discussion

The model offers a comprehensive and systematic description of the process of light signal transduction in guard cells and its crosstalk with CO_2 and ABA. The network representation we employ reveals the regulatory connections between seemingly remote components. For example, a recent publication, which was not included during the construction of the model, studied the indirect relationship between the SLAC1 anion channel and the regulation of K^+ uptake [22]. It reported that anion accumulation in the *slac1* anion channel knockout mutant induced the hyperpolarization of the plasma membrane which in turn promoted Ca^{2+} influx. Ca^{2+} influx led to an increase in the free cytosolic Ca^{2+} concentration, which then downregulated the inward K^+ channels. This relationship is supported by a path in the network, namely

AnionCh→PMV⊣CaIC→$[Ca^{2+}]_c$⊣K_{in}. Structural analysis of the network provided significant biological insight. The node degree offers a measure of node importance (Table 3-1) and path analysis reveals the robustness of the crosstalk between ABA and blue light in regulating stomatal opening. The strongly connected component analysis identified key components that mediate the cross-talk between blue light and ABA, such as $[Ca^{2+}]_c$ (Figure 3-2C), and highlighted the absence of cross-talk between red light and ABA (Figure 3-2A). The latter observation, recapitulated by the dynamic model, revealed the absence of experimental investigations of regulatory effects of ABA on red light-induced stomatal opening. Experiments performed in this study fill this knowledge gap and reveal that ABA does in fact inhibit red light-induced stomatal opening. We formulate the novel hypothesis that ABA inhibits sucrose accumulation, and demonstrate that integration of this hypothesis into the model restores the agreement between model and experiments.

The model reveals additional questions where further experimental investigation would be especially fruitful. I discuss a few such examples below. The model can offer a prediction for outcomes based on integration of the current knowledge, but it is up to future experiments to answer these questions definitively.

There has been a long debate on whether the plasma membrane H^+-ATPase is active under red light alone. The evidence regarding the status of the H^+-ATPase during red light-induced stomatal opening is mixed: activation of the proton pump by red light has been observed [78] or inferred [40], but the results in [78] were not reproduced [64,79] and the experiment in [40] was done under reduced CO_2 concentration. Little or no activation of the proton pump by red light has been observed in other experiments [18,80]. The model predicts the inactivity of the H^+-ATPase under red light in ambient air and predicts a moderate H^+-ATPase activity under red light in a

reduced CO_2 condition (see Figure 3-5B). Experiments dedicated to measuring the H^+-ATPase activity under red light with varying CO_2 concentrations will greatly improve the understanding of this matter.

A remaining question about sucrose as an osmoticum is the relative contribution of different sources of guard cell sucrose accumulation under different light and CO_2 conditions. Sugar accumulation, predominantly sucrose [35,81], can in theory result from photosynthetic carbon fixation, degradation of stored starch [6,81,82], or import from the apoplast [34,83,84]. These three processes exhibit different responsiveness to light and ABA. Photosynthesis is activated by either blue or red light, requires CO_2, and was not observed to be inhibited by ABA [47] (see Figure 3-2). Starch content was shown to be constant under red light but decreasing in time under blue light [6]. The H^+/sucrose symporter function requires apoplast acidification by the H^+-ATPase [34,83], which is not effective under red light in ambient air. Since ABA inhibits the plasma membrane H^+-ATPase [51,63,85] and apoplast acidification by the H^+-ATPase is required by the H^+/sucrose symporter, it can be inferred that the symporter activity can be inhibited by ABA. In the model I assumed that carbon fixation is the primary source of sucrose accumulation (see Supporting Text S1 of [106] for a detailed justification). Experiments that dissect the contribution of each source of guard cell sucrose accumulation will not only help improve the model, but also provide insight into the interaction between blue and red light and between light and ABA in the regulation of stomatal movement.

The simulations showed that the inactivation of photophosphorylation, e.g. by DCMU, induces a significant reduction in stomatal opening under all three light quality conditions (Table 3-2A). The model predicts that the inactivation of photophosphorylation i) reduces carbon fixation and hence the amount of sucrose accumulated via photosynthesis, and ii) reduces the amount of

ATP available for H^+-ATPase activity. Since the H^+-ATPase is not activated by red light in ambient air in the model, the simulations suggest that DCMU inhibits red light-induced stomatal opening through mechanism i) only. Since the H^+-ATPase is activated by blue light in ambient air, the simulations suggest that DCMU inhibits blue light-induced stomatal opening through both mechanism i) and ii). Experiments showed that DCMU partially inhibited blue light-induced stomatal opening and it completely inhibited red light-induced stomatal opening [3,80]; it would be informative to investigate the effect of DCMU on dual beam- or white light-induced stomatal opening as well. Further, it would also be interesting to explore the effect of DCMU and the respiratory inhibitor potassium cyanide (KCN) on light-induced stomatal opening in CO_2-free air, a condition under which photosynthetic carbon fixation is absent, as CO_2, the substrate for carbon fixation, is unavailable.

The model implements a brake-like effect of high C_i on H^+-ATPase activity based on the observation that an enhanced level of CO_2 depolarizes the plasma membrane [86] and the consequent hypothesis that elevated CO_2 inhibits the proton pump at the plasma membrane. This inhibitory effect of C_i on the H^+-ATPase activity helps to explain the activation of the H^+-ATPase under low CO_2 conditions [40]. Understanding the mechanism underlying this effect of C_i on the H^+-ATPase and in particular, whether it is a direct or an indirect effect, would provide valuable information. Such data would not only clarify the relationship between light (especially red light) and the H^+-ATPase activity level, but also offer insight into the synergy between blue and red light. According to the model, red light as a background to blue light can not only provide additional ATP through photophosphorylation for the H^+-ATPase activity, but also lower C_i via stimulation of mesophyll photosynthesis and thus raise the activity of the H^+-ATPase (Figure 3-5B). These two mechanisms could be critical in explaining the synergy between blue and red light in the intact

leaf. The model also predicts that ion uptake/accumulation, which hinges upon H^+-ATPase activity, is the primary mechanism for stomatal opening in response to red light under CO_2-free air (Table 3-3). Therefore, stomatal opening of a K_{in} knockout phenotype in response to red light with CO_2-free air should be severely impaired, in contrast to the minimal effect of K_{in} knockout on red light-induced stomatal opening under normal air (Table 3-4 and [73]). Experimental verification of this prediction would also support the model's predictions regarding the osmotic composition during stomatal opening in CO_2-free air (Table 3-3), and provide further evidence for the activation of the H^+-ATPase by red light in CO_2-free air as proposed in [40].

The current model offers a qualitatively accurate and quantitatively close depiction of short term stomatal movement in response to a light signal. There have been investigations, however, which demonstrate that under natural conditions (white light) sucrose accumulates in guard cells in the afternoon and replaces K^+ as the dominant osmoticum to maintain stomatal apertures [6]. An interesting potential future direction for the model is the incorporation of emerging knowledge concerning the cross-talk of guard cell circadian rhythms, light and ABA responses (e.g. [87]). The recent successful Boolean model of circadian clocks [88] makes the construction of such an integrated gene regulatory and signal transduction network model feasible.

The modeling framework characterizes each component with two or more levels and expresses the relationships between components as a mixture of logical rules and algebraic operations. Thus the model offers a parsimonious, computationally efficient yet quantitative description of the system's dynamics, making it a step forward from traditional Boolean models and an enhanced modeling tool for systems biology. The choices of the transfer functions of the nodes are a simplified and abstracted representation of the best available knowledge. Assuming discrete node levels is an approximation of reality (e.g. the concentration of a substance or the

potential across the plasma membrane is continuous in reality); there is, however, ample evidence of nonlinear regulation wherein not the concentration but rather its relationship with certain thresholds matters. Network-based discrete dynamic modeling has been successfully applied in a great variety of biological systems (reviewed in [89-92]). These models enabled the understanding of the systems and generated insightful predictions that were subsequently validated experimentally; recent examples include [24,93-97].

In the absence of detailed knowledge on relative reaction speeds I deliberately sampled different timescales by implementing random order asynchronous update. While this is not a fully accurate representation of reality, the averaged results of a large number of replicate simulations are representative of behaviors that are not sensitive to small changes in kinetic rates. Future observations of relative temporal patterns of multiple components or measurements of time delays among components can be incorporated by imposing restrictions on the update sequence (e.g. updating one group of nodes before another group [91]).

Having established the biological validity of the model, follow-up work in several directions is now possible, linking to recent advances in discrete and continuous dynamic modeling. Translating the model into a polynomial discrete dynamic system [98] or logical discrete model [99] would allow the use of software tools such as ADAM [100] or GINsim [99], and may yield further insights into the dynamic repertoire of the system. The model could also be translated into a Boolean model of an expanded network, where multi-level components are represented by multiple nodes in such a way that the group of binary nodes representing the same component allows the recapitulation of the same number of relative outcomes as the original multi-level node (see e.g. [97]). This transformation would allow the application of Boolean network analyses such as elementary signaling mode analysis and attractor analysis [91]. Through network simplification

methods [30,101], a core network with fewer nodes and edges could be distilled, which may be amenable for continuous modeling wherein differential equations replace the transfer functions. For instance, a model that integrates a simplified version of the model with OnGuard [20] would include the various signal transduction pathways, their cross-talk, and the quantitative description of ion flows in guard cells.

The model is generically adaptable, allowing one to incorporate emerging new pieces of information with ease. This modeling methodology can be readily applied to other systems where interaction and relative state information is available. The number of multi-level nodes can be minimized by identifying the node(s) for which more than three relative outcomes have been observed (indicative of a need for more than two levels), tracing upstream in the network, and inferring a minimal set of nodes whose multi-level nature could cause all the other nodes' multiple levels. The states of this set of nodes can then be defined as fundamental states (see Supporting Text S1 of [106]), and the states of other nodes subsequently derived through their updating rules. As demonstrated here, network-based dynamic models of biological systems can serve as a virtual control to test the coherence between experimental results generated in separate experiments, to generate predictions that inform and help prioritize future experiments, and to reveal new questions that deserve attention.

3.4 Materials and Methods

3.4.1 Constructing transfer functions for binary nodes

When there is only one upstream regulator of the target, the "*Equal*" rule is used for positive regulation and the "*Not*" rule is used for negative regulation. The "*AND*" operator is used when multiple regulators are required to activate the target. If each regulator is able to

independently activate the target, they are connected with the "*OR*" operator. For inhibition, the "*AND NOT*" operator is used, thereby requiring a low level or inactivity of the inhibitor in order for the target node to activate. One example of a Boolean transfer function is shown below:

$phot1_{complex}^{*}$ = phot1 *And* 14-3-3 $protein_{phot1}$

Phototropin 1 binds reversibly to a 14-3-3 protein (14-3-3 $protein_{phot1}$) upon the auto-phosphorylation of phototropin 1 in guard cells [102,103]. The 14-3-3 protein-phototropin complex ($phot1_{complex}$) is thought to confer an active state to phototropin 1, which then transmits the light signal to downstream elements. The *And* function connects phot1 and the 14-3-3 protein, indicating that the formation of the complex requires both of them.

3.4.2 Constructing transfer functions for nodes with more than two states

The transfer functions combine Boolean clauses with addition, subtraction, and multiplication. One example of a complex transfer function is given below:

PMV^{*} = PMV – H^{+}-$ATPase_{complex}$ + (AnionCh *And* (PMV = –2)) + (($[Ca^{2+}]_c$ = 2) *Or* KEV)

PMV is the difference of electric potential across the plasma membrane, i.e. the "membrane potential". Computationally, I use the value 0 to represent the resting potential of the plasma membrane. Negative values denote the further hyperpolarization of the plasma membrane, and positive values denote depolarization. I assume that five levels (-2, -1, 0, 1, 2) are sufficient for a qualitative description, and require that the PMV value stays bounded, i.e. the value will not further decrease (or increase) when it reaches -2 (or 2). The future PMV value (PMV^{*}) can be shifted away from or stay the same as its current value (PMV), depending on the hyperpolarizing and

depolarizing forces. Factors that cause hyperpolarization will decrease the PMV value (e.g. to -1 from 0), and factors that cause depolarization will increase the PMV value (e.g. from 0 to 1). An active 14-3-3 protein-bound H^+-ATPase causes the extrusion of H^+ from guard cell cytosol to the apoplast, hyperpolarizing the plasma membrane. Anion efflux at the plasma membrane causes the plasma membrane to depolarize. A steady anion efflux requires AnionCh to be active and it also requires plasma membrane depolarization. To avoid an oscillation in anion efflux and PMV induced by the discrete nature of the model, I require that PMV be -2 (most hyperpolarized) for effective anion efflux. A high $[Ca^{2+}]_c$ concentration (value 2) or K^+ release into the cytosol from vacuole (KEV) are modeled as independent factors causing the plasma membrane to depolarize. Supporting Text S5 of [106]indicates the pseudo-code for the implementation of two representative transfer functions.

3.4.3 Computational implementation and tools

The network in Figure 3-1 was drawn with the software yED (http://www.yworks.com/en/products_yed_about.html). The network analyses (strongly connected component identification, calculation of the number of simple paths between two nodes) were implemented by custom MATLAB code. Similar analyses can also be done with one of the following tools: NetworkX, a Python graph software library [104], Cytoscape, a network integration, visualization and analysis tool [105], or MATLAB's graph theory toolbox, grTheory.

The dynamic model was implemented by custom MATLAB code. The pseudo-code of the simulations is indicated in Supporting Text S5 of [106].

3.5 References

1. Allaway WG, Mansfield TA (1967) Stomatal Responses to Changes in Carbon Dioxide Concentration in Leaves Treated with 3-(4-Chlorophenyl)-I, I-Dimethylurea. New Phytologist 66: 7.
2. Gepstein S, Jacobs M, Taiz L (1982) Inhibition of stomatal opening in Vicia faba epidermal tissue by vanadate and abscisic acid. Plant Science Letters 28: 10.
3. Schwartz A, Zeiger E (1984) Metabolic energy for stomatal opening. Roles of photophosphorylation and oxidative phosphorylation. Planta 161: 129-136.
4. Assmann SM, Simoncini L, Schroeder JI (1985) Blue light activates electrogenic ion pumping in guard cell protoplasts of Vicia faba L. Nature 318: 3.
5. Assmann SM (1988) Enhancement of the Stomatal Response to Blue Light by Red Light, Reduced Intercellular Concentrations of CO(2), and Low Vapor Pressure Differences. Plant physiology 87: 226-231.
6. Tallman G, Zeiger E (1988) Light quality and osmoregulation in vicia guard cells : evidence for involvement of three metabolic pathways. Plant physiology 88: 887-895.
7. Schroeder JI, Allen GJ, Hugouvieux V, Kwak JM, Waner D (2001) Guard Cell Signal Transduction. Annual review of plant physiology and plant molecular biology 52: 627-658.
8. Roelfsema MR, Hedrich R (2005) In the light of stomatal opening: new insights into 'the Watergate'. The New phytologist 167: 665-691.
9. Shimazaki K, Doi M, Assmann SM, Kinoshita T (2007) Light regulation of stomatal movement. Annual review of plant biology 58: 219-247.
10. Assmann SM, Shimazaki K (1999) The multisensory guard cell. Stomatal responses to blue light and abscisic acid. Plant physiology 119: 809-816.
11. Mott KA (1988) Do Stomata Respond to CO(2) Concentrations Other than Intercellular? Plant physiology 86: 200-203.
12. Assmann SM, Grantz DA (1990) Stomatal response to humidity in sugarcane and soybean: effect of vapor pressure difference on the kinetics of the blue light response. Plant Cell Environment 13: 7.
13. Ogawa T, Ishikawa H, Shimada K, Shibata K (1978) Synergistic action of red and blue light and action spectra for malate formation in guard cells of Vicia faba L. Planta 142: 61-65.
14. Huala E, Oeller PW, Liscum E, Han IS, Larsen E, et al. (1997) Arabidopsis NPH1: a protein kinase with a putative redox-sensing domain. Science 278: 2120-2123.
15. Briggs WR, Beck CF, Cashmore AR, Christie JM, Hughes J, et al. (2001) The phototropin family of photoreceptors. Plant Cell 13: 993-997.
16. Kinoshita T, Doi M, Suetsugu N, Kagawa T, Wada M, et al. (2001) Phot1 and phot2 mediate blue light regulation of stomatal opening. Nature 414: 656-660.
17. Takemiya A, Shimazaki K (2010) Phosphatidic acid inhibits blue light-induced stomatal opening via inhibition of protein phosphatase 1 [corrected]. Plant physiology 153: 1555-1562.
18. Hayashi M, Inoue S, Takahashi K, Kinoshita T (2011) Immunohistochemical detection of blue light-induced phosphorylation of the plasma membrane H(+)-ATPase in stomatal guard cells. Plant & cell physiology 52: 1238-1248.
19. Li S, Assmann SM, Albert R (2006) Predicting essential components of signal transduction networks: a dynamic model of guard cell abscisic acid signaling. PLoS biology 4: e312.
20. Hills A, Chen ZH, Amtmann A, Blatt MR, Lew VL (2012) OnGuard, a computational platform for quantitative kinetic modeling of guard cell physiology. Plant physiology 159: 1026-1042.

21. Chen ZH, Hills A, Batz U, Amtmann A, Lew VL, et al. (2012) Systems dynamic modeling of the stomatal guard cell predicts emergent behaviors in transport, signaling, and volume control. Plant physiology 159: 1235-1251.
22. Wang Y, Papanatsiou M, Eisenach C, Karnik R, Williams M, et al. (2012) Systems dynamic modeling of a guard cell Cl- channel mutant uncovers an emergent homeostatic network regulating stomatal transpiration. Plant physiology 160: 1956-1967.
23. Naldi A, Carneiro J, Chaouiya C, Thieffry D (2010) Diversity and plasticity of Th cell types predicted from regulatory network modelling. PLoS computational biology 6: e1000912.
24. Saez-Rodriguez J, Simeoni L, Lindquist JA, Hemenway R, Bommhardt U, et al. (2007) A logical model provides insights into T cell receptor signaling. PLoS computational biology 3: e163.
25. Fumia HF, Martins ML (2013) Boolean network model for cancer pathways: predicting carcinogenesis and targeted therapy outcomes. PloS one 8: e69008.
26. Helikar T, Kowal B, Madrahimov A, Shrestha M, Pedersen J, et al. (2012) Bio-logic builder: a non-technical tool for building dynamical, qualitative models. PloS one 7: e46417.
27. Kachalo S, Zhang R, Sontag E, Albert R, DasGupta B (2008) NET-SYNTHESIS: a software for synthesis, inference and simplification of signal transduction networks. Bioinformatics 24: 293-295.
28. Blatt MR (2000) Cellular signaling and volume control in stomatal movements in plants. Annual review of cell and developmental biology 16: 221-241.
29. Hayashi M, Kinoshita T (2011) Crosstalk between blue-light- and ABA-signaling pathways in stomatal guard cells. Plant signaling & behavior 6: 1662-1664.
30. Saadatpour A, Albert R, Reluga T (2013) A reduction method for Boolean network models proven to conserve attractors. SIAM Journal on Applied Dynamical Systems 12: 15.
31. Garcia-Mata C, Gay R, Sokolovski S, Hills A, Lamattina L, et al. (2003) Nitric oxide regulates K(+) and Cl(-) channels in guard cells through a subset of abscisic acid-evoked signaling pathways. Proceedings of the National Academy of Sciences of the United States of America 100: 11116-11121.
32. Joudoi T, Shichiri Y, Kamizono N, Akaike T, Sawa T, et al. (2013) Nitrated cyclic GMP modulates guard cell signaling in Arabidopsis. Plant Cell 25: 558-571.
33. Willmer CM, Fricker MD (1996) Stomata. London, UK: Chapman & Hall.
34. Reddy AR, Das VSR (1986) Stomatal movements and sucrose uptake by guard cell protoplasts of Commelina benghalensis L. Plant & cell physiology 27: 6.
35. Talbott LD, Zeiger E (1993) Sugar and Organic Acid Accumulation in Guard Cells of Vicia faba in Response to Red and Blue Light. Plant physiology 102: 1163-1169.
36. Kinoshita T, Shimazaki K (1999) Blue light activates the plasma membrane H(+)-ATPase by phosphorylation of the C-terminus in stomatal guard cells. The EMBO journal 18: 5548-5558.
37. Raschke K, Humble GD (1973) No uptake of anions required by opening stomata of Vicia faba: Guard cells release hydrogen ions. Planta 115: 47-57.
38. Blatt MR (1992) K(+) channels of stomatal guard cells. Characteristics of the inward rectifier and its control by pH. The Journal of general physiology 99: 615-644.
39. Hsiao TC, Allaway WG (1973) Action Spectra for Guard Cell Rb Uptake and Stomatal Opening in Vivia faba. Plant physiology 51: 82-88.

40. Olsen RL, Pratt RB, Gump P, Kemper A, Tallman G (2002) Red light activates a chloroplast-dependent ion uptake mechanism for stomatal opening under reduced CO2 concentrations in Vicia spp. New Phytologist 153: 12.
41. Schwartz A, Ilan N, Schwarz M, Scheaffer J, Assmann SM, et al. (1995) Anion-Channel Blockers Inhibit S-Type Anion Channels and Abscisic Acid Responses in Guard Cells. Plant physiology 109: 651-658.
42. Jacob T, Ritchie S, Assmann SM, Gilroy S (1999) Abscisic acid signal transduction in guard cells is mediated by phospholipase D activity. Proceedings of the National Academy of Sciences of the United States of America 96: 12192-12197.
43. Staxen I, Pical C, Montgomery LT, Gray JE, Hetherington AM, et al. (1999) Abscisic acid induces oscillations in guard-cell cytosolic free calcium that involve phosphoinositide-specific phospholipase C. Proceedings of the National Academy of Sciences of the United States of America 96: 1779-1784.
44. Zhang X, Wang H, Takemiya A, Song CP, Kinoshita T, et al. (2004) Inhibition of blue light-dependent H(+) pumping by abscisic acid through hydrogen peroxide-induced dephosphorylation of the plasma membrane H(+)-ATPase in guard cell protoplasts. Plant physiology 136: 4150-4158.
45. Roelfsema MR, Hedrich R, Geiger D (2012) Anion channels: master switches of stress responses. Trends in plant science 17: 221-229.
46. Barabasi AL, Oltvai ZN (2004) Network biology: understanding the cell's functional organization. Nature reviews Genetics 5: 101-113.
47. Mawson BT, Colman B, Cummins WR (1981) Abscisic Acid and photosynthesis in isolated leaf mesophyll cell. Plant physiology 67: 233-236.
48. Schwartz A (1985) Role of Ca and EGTA on Stomatal Movements in Commelina communis L. Plant physiology 79: 1003-1005.
49. McAinsh MR, Brownlee C, Hetherington AM (1990) Abscisic acid-induced elevation of guard cell cytosolic Ca(2+) precedes stomatal closure. Nature 343: 3.
50. Shimazaki K, Kinoshita T, Nishimura M (1992) Involvement of Calmodulin and Calmodulin-Dependent Myosin Light Chain Kinase in Blue Light-Dependent H Pumping by Guard Cell Protoplasts from Vicia faba L. Plant physiology 99: 1416-1421.
51. Roelfsema MR, Staal M, Prins HBA (1998) Blue light-induced apoplastic acidification of Arabidopsis thaliana guard cells: inhibition by ABA is mediated through protein phosphatases. Physiologia Plantarum 103: 9.
52. Baum G, Long JC, Jenkins GI, Trewavas AJ (1999) Stimulation of the blue light phototropic receptor NPH1 causes a transient increase in cytosolic Ca(2+). Proceedings of the National Academy of Sciences of the United States of America 96: 13554-13559.
53. Harada A, Shimazaki K (2009) Measurement of changes in cytosolic Ca(2+) in Arabidopsis guard cells and mesophyll cells in response to blue light. Plant & cell physiology 50: 360-373.
54. Pei ZM, Ward JM, Harper JF, Schroeder JI (1996) A novel chloride channel in Vicia faba guard cell vacuoles activated by the serine/threonine kinase, CDPK. The EMBO journal 15: 6564-6574.
55. Schroeder JI, Hagiwara S (1989) Cytosolic calcium regulates ion channels in the plasma membrane of Vicia faba guard cells. Nature 338: 4.
56. Kinoshita T, Nishimura M, Shimazaki K (1995) Cytosolic Concentration of Ca(2+) Regulates the Plasma Membrane H(+)-ATPase in Guard Cells of Fava Bean. Plant Cell 7: 1333-1342.

57. Iino M, Ogawa T, Zeiger E (1985) Kinetic properties of the blue-light response of stomata. Proceedings of the National Academy of Sciences of the United States of America 82: 8019-8023.
58. Karlsson PE (1986) Blue light regulation of stomata in wheat seedlings. I. Influence of red background illumination and initial conductance level. Physiologia Plantarum 66: 5.
59. Sharkey TD, Ogawa T (1987) Stomatal responses to light. In: Zeiger E, Farquhar G, Cowan I, editors. Stomatal Function. Stanford, CA: Stanford Univ. Press. pp. 195-208.
60. Shimazaki K, Goh CH, Kinoshita T (1999) Involvement of intracellular Ca(2+) in blue light-dependent proton pumping in guard cell protoplasts from Vicia faba. Physiologia Plantarum 105: 8.
61. McAinsh MR, Webb A, Taylor JE, Hetherington AM (1995) Stimulus-Induced Oscillations in Guard Cell Cytosolic Free Calcium. Plant Cell 7: 1207-1219.
62. Chaves M, Albert R, Sontag ED (2005) Robustness and fragility of Boolean models for genetic regulatory networks. Journal of theoretical biology 235: 431-449.
63. Shimazaki K, Iino M, Zeiger E (1986) Blue light-dependent proton extrusion by guard-cell protoplasts of Vicia faba. Nature 319: 3.
64. Taylor AR, Assmann SM (2001) Apparent absence of a redox requirement for blue light activation of pump current in broad bean guard cells. Plant physiology 125: 329-338.
65. Kinoshita T, Shimazaki K (2002) Biochemical evidence for the requirement of 14-3-3 protein binding in activation of the guard-cell plasma membrane H(+)-ATPase by blue light. Plant & cell physiology 43: 1359-1365.
66. Turner NC (1972) K(+) uptake of guard cells stimulated by fusicoccin. Nature 235: 2.
67. Squire GR, Mansfield TA (1974) The action of fusicoccin on stomatal guard cells and the subsidiary cells. New Phytologist 73: 8.
68. Marrè E (1979) Fusicoccin: A tool in plant physiology. Annual Review of Plant Physiology 30: 16.
69. Johansson F, Sommarin M, Larsson C (1993) Fusicoccin Activates the Plasma Membrane H(+)-ATPase by a Mechanism Involving the C-Terminal Inhibitory Domain. Plant Cell 5: 321-327.
70. Palmgren MG (2001) PLANT PLASMA MEMBRANE H(+)-ATPases: Powerhouses for Nutrient Uptake. Annual review of plant physiology and plant molecular biology 52: 817-845.
71. Schwartz A, Illan N, Assmann SM (1991) Vanadate inhibition of stomatal opening in epidermal peels of Commelina communis: Cl(-) interferes with vanadate uptake. Planta 183: 590-596.
72. Amodeo G, Srivastava A, Zeiger E (1992) Vanadate inhibits blue light-stimulated swelling of vicia guard cell protoplasts. Plant physiology 100: 1567-1570.
73. Lebaudy A, Vavasseur A, Hosy E, Dreyer I, Leonhardt N, et al. (2008) Plant adaptation to fluctuating environment and biomass production are strongly dependent on guard cell potassium channels. Proceedings of the National Academy of Sciences of the United States of America 105: 5271-5276.
74. Jeon BW, Hwang JU, Hwang Y, Song WY, Fu Y, et al. (2008) The Arabidopsis small G protein ROP2 is activated by light in guard cells and inhibits light-induced stomatal opening. Plant Cell 20: 75-87.

75. Hamilton DW, Hills A, Kohler B, Blatt MR (2000) Ca(2+) channels at the plasma membrane of stomatal guard cells are activated by hyperpolarization and abscisic acid. Proceedings of the National Academy of Sciences of the United States of America 97: 4967-4972.
76. Pei ZM, Murata Y, Benning G, Thomine S, Klusener B, et al. (2000) Calcium channels activated by hydrogen peroxide mediate abscisic acid signalling in guard cells. Nature 406: 731-734.
77. Suhita D, Raghavendra AS, Kwak JM, Vavasseur A (2004) Cytoplasmic alkalization precedes reactive oxygen species production during methyl jasmonate- and abscisic acid-induced stomatal closure. Plant physiology 134: 1536-1545.
78. Serrano EE, Zeiger E, Hagiwara S (1988) Red light stimulates an electrogenic proton pump in Vicia guard cell protoplasts. Proceedings of the National Academy of Sciences of the United States of America 85: 436-440.
79. Roelfsema MR, Steinmeyer R, Staal M, Hedrich R (2001) Single guard cell recordings in intact plants: light-induced hyperpolarization of the plasma membrane. The Plant journal : for cell and molecular biology 26: 1-13.
80. Tominaga M, Kinoshita T, Shimazaki K (2001) Guard-cell chloroplasts provide ATP required for H(+) pumping in the plasma membrane and stomatal opening. Plant & cell physiology 42: 795-802.
81. Talbott LD, Zeiger E (1998) The role of sucrose in guard cell osmoregulation. Journal of experimental botany 49: 9.
82. Outlaw WH, Manchester J (1979) Guard cell starch concentration quantitatively related to stomatal aperture. Plant physiology 64: 79-82.
83. Ritte G, Rosenfeld J, Rohrig K, Raschke K (1999) Rates of sugar uptake by guard cell protoplasts of pisum sativum L. Related To the solute requirement for stomatal opening. Plant physiology 121: 647-656.
84. Stadler R, Buttner M, Ache P, Hedrich R, Ivashikina N, et al. (2003) Diurnal and light-regulated expression of AtSTP1 in guard cells of Arabidopsis. Plant physiology 133: 528-537.
85. Goh CH, Kinoshita T, Oku T, Shimazaki K (1996) Inhibition of Blue Light-Dependent H(+) Pumping by Abscisic Acid in Vicia Guard-Cell Protoplasts. Plant physiology 111: 433-440.
86. Edwards A, Bowling DJF (1985) Evidence for a CO2 inhibited proton extrusion pump in the stomatal cells of Tradescantia virginiana. Journal of experimental botany 36: 8.
87. Pokhilko A, Mas P, Millar AJ (2013) Modelling the widespread effects of TOC1 signalling on the plant circadian clock and its outputs. BMC systems biology 7: 23.
88. Akman OE, Watterson S, Parton A, Binns N, Millar AJ, et al. (2012) Digital clocks: simple Boolean models can quantitatively describe circadian systems. Journal of the Royal Society, Interface / the Royal Society 9: 2365-2382.
89. Albert R, Assmann SM (2009) An Overview of System Biology. Annual Plant Reviews 35: 26.
90. Morris MK, Saez-Rodriguez J, Sorger PK, Lauffenburger DA (2010) Logic-based models for the analysis of cell signaling networks. Biochemistry 49: 3216-3224.
91. Wang RS, Saadatpour A, Albert R (2012) Boolean modeling in systems biology: an overview of methodology and applications. Physical biology 9: 055001.

92. Wynn ML, Consul N, Merajver SD, Schnell S (2012) Logic-based models in systems biology: a predictive and parameter-free network analysis method. Integrative biology : quantitative biosciences from nano to macro 4: 1323-1337.
93. Thakar J, Pilione M, Kirimanjeswara G, Harvill ET, Albert R (2007) Modeling systems-level regulation of host immune responses. PLoS computational biology 3: e109.
94. Zhang R, Shah MV, Yang J, Nyland SB, Liu X, et al. (2008) Network model of survival signaling in large granular lymphocyte leukemia. Proceedings of the National Academy of Sciences of the United States of America 105: 16308-16313.
95. Saadatpour A, Wang RS, Liao A, Liu X, Loughran TP, et al. (2011) Dynamical and structural analysis of a T cell survival network identifies novel candidate therapeutic targets for large granular lymphocyte leukemia. PLoS computational biology 7: e1002267.
96. Thakar J, Pathak AK, Murphy L, Albert R, Cattadori IM (2012) Network model of immune responses reveals key effectors to single and co-infection dynamics by a respiratory bacterium and a gastrointestinal helminth. PLoS computational biology 8: e1002345.
97. Miskov-Zivanov N, Turner MS, Kane LP, Morel PA, Faeder JR (2013) The duration of T cell stimulation is a critical determinant of cell fate and plasticity. Science signaling 6: ra97.
98. Veliz-Cuba A, Jarrah AS, Laubenbacher R (2010) Polynomial algebra of discrete models in systems biology. Bioinformatics 26: 1637-1643.
99. Naldi A, Berenguier D, Faure A, Lopez F, Thieffry D, et al. (2009) Logical modelling of regulatory networks with GINsim 2.3. Bio Systems 97: 134-139.
100. Hinkelmann F, Brandon M, Guang B, McNeill R, Blekherman G, et al. (2011) ADAM: analysis of discrete models of biological systems using computer algebra. BMC bioinformatics 12: 295.
101. Berenguier D, Chaouiya C, Monteiro PT, Naldi A, Remy E, et al. (2013) Dynamical modeling and analysis of large cellular regulatory networks. Chaos 23: 025114.
102. Kinoshita T, Emi T, Tominaga M, Sakamoto K, Shigenaga A, et al. (2003) Blue-light- and phosphorylation-dependent binding of a 14-3-3 protein to phototropins in stomatal guard cells of broad bean. Plant physiology 133: 1453-1463.
103. Sullivan S, Thomson CE, Kaiserli E, Christie JM (2009) Interaction specificity of Arabidopsis 14-3-3 proteins with phototropin receptor kinases. FEBS letters 583: 2187-2193.
104. Hagberg AA, Schult DA, Swart PJ (2008) Exploring network structure, dynamics, and function using NetworkX. Proceedings of the 7th Python in Science Conference (SciPy2008): 11-15.
105. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, et al. (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Research 13: 2498-2504.
106. Sun Z, Jin X, Albert R, Assmann SM. (2014) Multi-level modeling of light-induced stomatal opening offers new insights into its regulation by drought. PLoS computational biology 10:e1003930.

Chapter 4

Node-independent elementary signaling modes: a measure of redundancy in Boolean signaling transduction networks

4.1 Introduction

The explanatory and predictive power of integrating the components of a biological system and their interactions into a network model has been demonstrated over the past few decades [1-5]. Network models that give insight into the functional outcomes of the relevant biological network (for example, by connecting molecular interactions with cellular behaviors like motion or cell death) are especially useful. A promising method toward this functional integration is the construction of so-called expanded networks or hyper-graphs that incorporate not only the edges incident on each node but also the coordination among these edges [6-9]. Specifically, a Boolean function (using the logic operators NOT, AND, OR) is constructed for each node (see Figure 4-1A), and then represented by, for example, the addition of complementary nodes (which reflect NOT relationships) and composite nodes (which reflect AND relationships, see Figure 4-1B). Indeed, in the spectrum of different dynamic models, Boolean models, successfully used in a wide range of biological domains, are the most parsimonious [10]. This parsimony is combined with the power of providing mechanistic insights, predicting system behaviors and offering guidance for future experiments.

In this study, we focus on signal transduction networks, which we define generally as directed networks that have one or a few input (source) nodes and one or a few output (sink) nodes. The expanded network has been used to determine the number of pathways that can independently carry out signal transduction from an input node to an output node, to quantify the importance of

each node to this signal transduction process [6], and to predict the network's long-term behaviors [7]. Here we introduce the concept of node-independent elementary signaling modes, which relate to a signal transduction network's functional redundancy. We present a rigorous definition and the procedure to identify the maximal set of node-independent elementary signaling modes of a Boolean network. We analyze both theoretical and empirical network examples and discuss the results' interpretation and implications for real networks.

4.2 Results

4.2.1 Independent elementary signaling modes and functional path redundancy

The concept of an elementary signaling mode (ESM) between an input (source node) and an output (sink node) of a biological signal transduction network, or any directed network, was introduced in [6]. An ESM describes a minimal set of nodes that can carry out the signal transduction from the input to the output. It can be considered as a 'unit' for the successful performance of signal transduction: all nodes of an ESM being functional is a necessary and sufficient condition for the signal to propagate from the input to the output regardless of the status of any node outside of that ESM. The shortest path between the input and output, another commonly used network concept, is contained within at least one ESM. Indeed an ESM may be a path from the input to the output (see Figure 4-1C). In many cases, however, the activation of one or more nodes on an input-output path requires their regulation by nodes outside of the path; in these cases the relevant ESM will contain the required regulators as well.

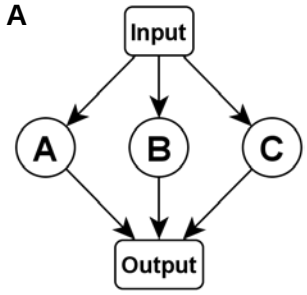
Separate ESMs of a network might share nodes and/or edges. In other words, a node or an edge may be contained in one or more ESMs of the network. To eliminate this ambiguity, we introduce the concept of node-independent ESMs: two ESMs that correspond to the same input and output node are node-independent if they do not share nodes other than the input and output

node. Intuitively, the number of node-independent ESMs characterizes the functional path redundancy of the input-output pair. Here we use 'redundant' to indicate that the node-independent ESMs are independent and functionally equivalent and thus they are interchangeable in case of errors, rendering the propagation of the signal robust.

The accumulation of experimental evidence makes increasingly possible the construction of network models with multiple input nodes and/or output nodes. The functional path redundancy of a signal transduction network is a composite of the functional path redundancy of its input-output pairs. Although the concept of node-independent ESMs and the methodology introduced here can be straightforwardly extended to include different input-output combinations, the functional redundancy of the signaling cascade triggered by a given input node and a given output node is the main focus of this work.

To quantify the path redundancy of an input-output pair, we define a *full set* of node-independent ESMs connecting the input to the output such that: (i) all ESMs within the set are pairwise node-independent, and (ii) any ESM that is not in the set shares at least one node with at least one ESM that is in the set. Different full sets of node-independent ESMs might exist for the same input-output pair. Among them, we call the set that has the highest number of ESMs the maximum set of node-independent ESMs of a signal transduction network. The number of ESMs within the maximum set then quantifies the functional path redundancy of the input-output pair.

4.2.2 Procedure to identify node-independent ESMs



Node	Boolean Transfer Function
Input	-
A	A*=Input
B	B*=Input
C	C*=Input
Output	Output*=A *Or* (B *And* C)

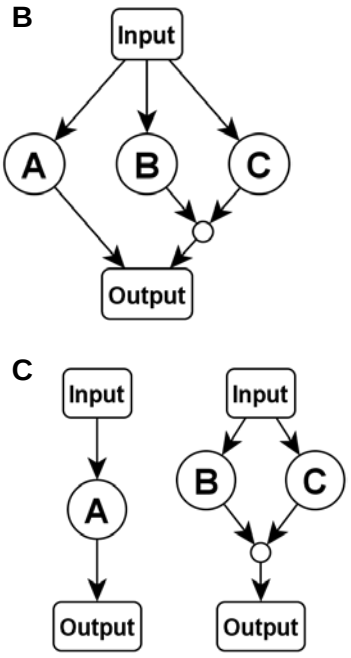


Figure 4-1. An illustration of a simple signal transduction network and a Boolean model of signal propagation. (A) The original network and the table of Boolean transfer functions describing the regulation of each node. Each node is characterized by a binary state. The Boolean transfer function indicates the future state of the node (marked by the asterisk) as a function of the current state of its regulators. For simplicity the node states are denoted by the node names. (B) The expanded network. A composite node (small unlabeled circle) is introduced to denote that node B and node C are connected with the Boolean function *And*. This way multiple edges targeting the same composite node are always connected by *And*, while multiple edges targeting the same real (non-composite) node are connected by *Or*. One no longer needs both the network structure and the table of Boolean transfer functions since the regulation is now incorporated into and explicitly represented by the expanded network. (C) The ESMs of the network. The network has two ESMs, and they are node-independent from each other.

To facilitate the identification of the nodes that are functionally related, the network needs to be first 'expanded' with the introduction of composite nodes to represent regulatory synergy between nodes (represented by the Boolean *And* operator) [6]. The expansion procedure used in this study follows [6], with the exception that inhibitory/negative edges are preserved in the expanded network. Upon the expansion of the original network, the complete list of all ESMs $\{ESM_1, ESM_2, \ldots, ESM_i, \ldots, ESM_j, \ldots, ESM_N\}$ can be obtained using a bottom-up subgraph

growing algorithm first described in [11]. Specifically, starting with a seed that contains only the output node, one makes as many copies of the output node as the number of edges that target the output, with each edge and upstream node separated into a different copy. The next step is to examine the upstream node in each copy: if the node is real (not composite), then as many copies are made as the number of incoming edges to that upstream node, and each incoming edge and the corresponding upstream regulators will be again separated into different copies; if the node is composite (i.e. representing a Boolean *And* relation), all incoming edges and their upstream effectors are included in the same copy. The process iterates until the upstream regulator found is either already in the copy, indicating a cycle, or it is the input node, where that branch of the search will end and the next node or subgraph copy in line will be examined.

It is straightforward to determine which ESMs share nodes by intersecting the list of nodes of each ESM. We propose a concise representation focusing on independence: we construct an ESM contingency graph whose nodes are ESMs. In this contingency graph an edge exists between ESM_i and ESM_j if ESM_i and ESM_j do not share any node (except of the input and output which are shared by all ESMs). All pairwise dependencies are checked to determine the presence or absence of edges between pairs of ESMs in the ESM contingency graph. The full sets of node-independent ESMs correspond to the maximum cliques (complete subgraphs) in the contingency graph, and the maximum set is defined by the maximal clique in the contingency graph. This is analogous to the related concept of a set of independent nodes in a graph, which is solved by determining the maximal cliques of the complementary graph, wherein nodes are connected by an edge if they are not connected by an edge in the original graph. Clique finding is a known NP-complete problem. Fortunately, we have found that the ESM contingency graph of many networks is sparse.

4.2.3 Theoretical examples of expanded networks, their ESMs and ESM contingency graphs

We exemplify node-independent ESMs through prototypical networks rich in network motifs [12].

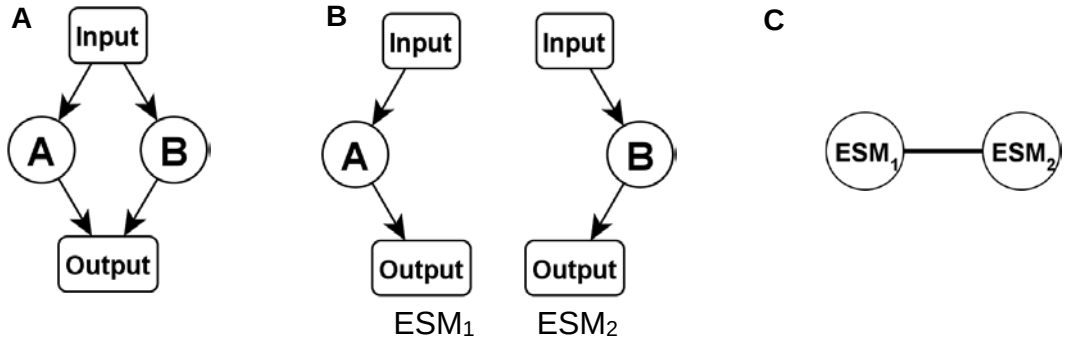


Figure 4-2. A bi-parallel signal transduction motif from the input to the output node contains two node-independent ESMs. (A) The expanded network. Both real (non-composite) nodes A and B have an edge directly incident upon the output node, implying that the Boolean transfer function for the output node is Output*=A *Or* B. (B) The two ESMs of the expanded network. (C) The ESM contingency graph. Each ESM is represented as a node. Since ESM_1 and ESM_2 are node-independent, their corresponding nodes are connected.

Figure 4-2 shows a bi-parallel signal transduction network in its simplest form. This network has two ESMs which are node-independent (Figure 4-2B). The corresponding ESM contingency graph has two nodes (for the two ESMs) connected by an edge (Figure 4-2C).

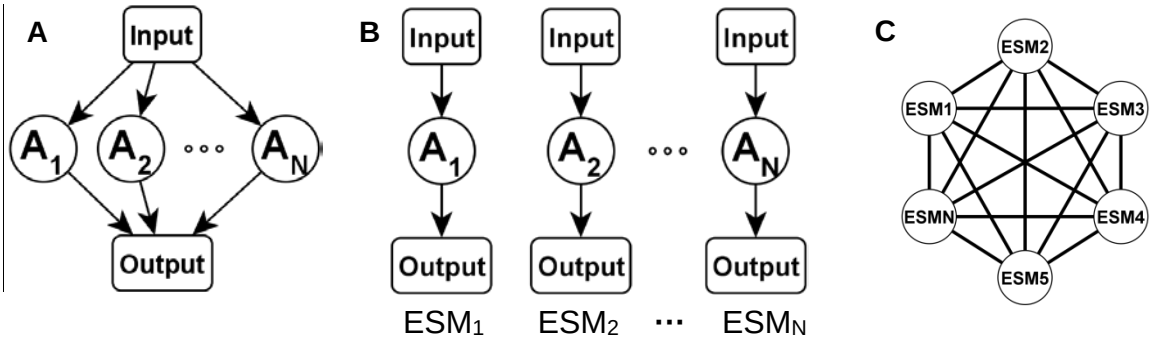
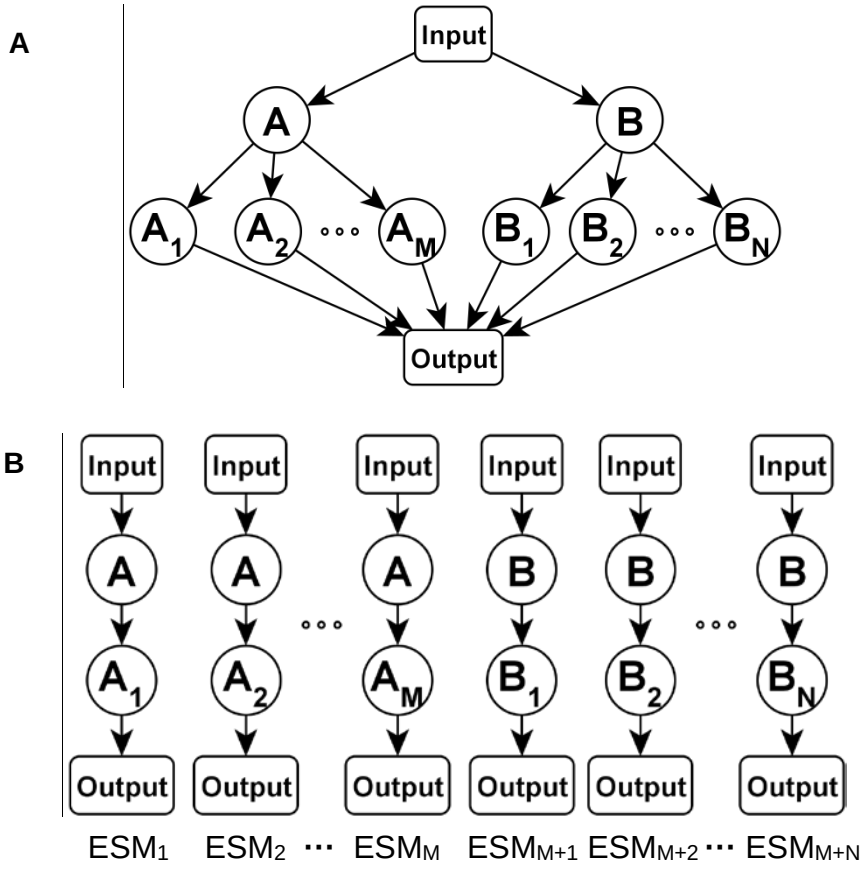


Figure 4-3. N parallel paths from the input to the output. The Boolean transfer function for the output node is Output*=A_1 *Or* A_2 *Or* … *Or* A_N. (A) The expanded network does not have any composite nodes. (B) There are N ESMs, all pair-wise node-independent. (C) The ESM contingency graph is a complete graph, as any pair of ESMs is connected by an edge.

Figure 4-3 illustrates the general case of N parallel paths from the input to the output. Every pair of node-independent ESMs is connected by an edge in the ESM contingency graph. A group of node-independent ESMs therefore forms a complete graph, or clique, in the ESM contingency graph (Figure 4-3C). If N=3, the ESM contingency graph forms a three-clique (triangle). Any maximal subgraph (i.e. a subset of the nodes and all the edges among them) of the clique is a clique, too. Therefore, any subset of the ESMs represented by the clique forms a set of node-independent ESMs as well, but the solution is not a full set since more ESMs can be added to the subset and the set would still be node-independent. To identify all full sets of node-independent ESMs of a network, one needs to find all maximal cliques in the ESM contingency graph, since each maximal clique corresponds to one solution.



C

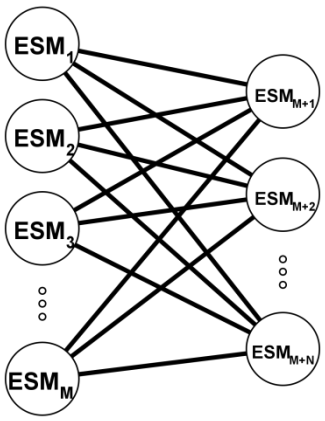


Figure 4-4. The formation of a bipartite structure in the ESM contingency graph and the network from which it arises. The Boolean transfer function for the output node is $Output^{*}=A_1$ *Or* A_2 *Or* … *Or* A_M *Or* B_1 *Or* B_2 *Or* … *Or* B_N. (A) The expanded network does not have any composite nodes. (B) The list of M+N ESMs. (C) The ESM contingency graph has a bipartite structure.

If the signal propagates from the input to the output through two distinct mechanisms, each of which is characterized by several ESMs that all share a node (e.g., node A, B, Figure 4-4), a bipartite structure in the ESM contingency graph arises. None of the ESMs that share node A are node-independent. The same argument holds for the mechanism signified by the common usage of node B. Any ESM in the first group, however, is node-independent from any ESM in the second group. The resulting ESM contingency graph henceforth has a bipartite structure. If M=1 or N=1, the bipartite contingency graph collapses into a tree structure. If M=N=2, the bipartite contingency graph becomes a cycle of 4 nodes. Regardless of the value of M and N, the bipartite structure implies that the largest clique has two nodes, thus the full set of node-independent ESMs has two members. If there are n independent mechanisms, the ESM contingency graph becomes n-partite, whose biggest clique has n nodes, leading to n ESMs in the full node-independent ESM set.

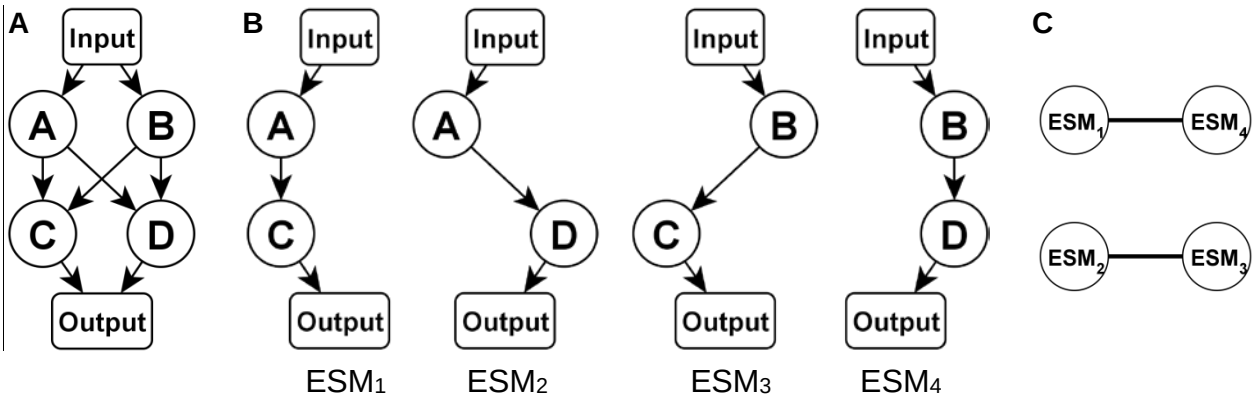


Figure 4-5. The bifan motif, its ESM list and its ESM contingency graph. The Boolean transfer function for the output node is Output*=C *Or* D. (A) The expanded network. (B) The list of ESMs. (C) The ESM contingency graph.

Lastly, we use what is called a bifan motif [13] to illustrate that the ESM contingency graph can be disconnected, and multiple sets of node-independent ESMs may exist. In this network (Figure 4-5A), C and D can be activated by either A or B, and the output can be activated by either C or D. ESM_1 and ESM_4 are node-independent, and ESM_2 and ESM_3 are node-independent. This is reflected by two two-cliques in the ESM contingency graph. There is no edge, however, between the two cliques. Both cliques are maximum sets of node-independent ESMs.

4.2.4 Identification of node-independent ESMs in real signaling networks

A rich repertoire of signaling network models for a variety of biological systems is now available. We applied our methodology to five discrete dynamic models of biological systems [14-18]. We found that despite the existence of numerous ESMs in each network, the number of ESMs in the maximum node-independent ESM sets for these systems is one [14-18]. This indicates the absence of more than one node-independent signal transduction mechanism between the input and the output, as one or more nodes are shared by some or all signal transduction mechanisms in these systems. In order to be able to analyze these cases further we introduce the concept of reusable nodes: nodes that can appear repeatedly across the list of ESMs. Node-independence of two ESMs

can then be relaxed to mean that no node other than the reusable nodes is shared by the two ESMs. In cases where there exist nodes that appear in all ESMs, they need to be included in the reusable set in order to have a full set of node-independent ESMs with more than one member. We call the set of nodes shared by all ESMs the mandatory reusable set. Table 4-1 indicates the analysis results of the five networks, including the mandatory reusable node set, whose sharing is a condition for having more than one node-independent ESM in each network.

Network	Input Node(s)	Output Node	Number of Nodes	Number of Edges	Number of ESMs	Maximum Number of Node-Independent ESMs	Mandatory Reusable Node Set
ABA Signaling [14]	ABA	Stomatal Closure	55	90	104	2	Depolar, AnionEM, Actin, malate, PEPC
Host Immune Response [15]	Bacteria	Phagocytosis	18	42	120	2	Epithelial cells, Complement, Ag-Ab complex, Pro-inflammatory cytokines, Recruited PMNs, Activated Phagocytic Cells, Macrophages, Th1 cells, T0 cells, Th2 cells, B cells, Th1 related cytokines, Th2 related cytokines, Dendritic cells
T-cell receptor signaling [16]	CD28, TCRlig	CRE	41	72	62	2	TCRb, ZAP70, DAG, SLP76, cCblp2, Vav1, Gab2, Abl, Fyn, ERK, PIP3, Itk, PLCga, TCRp, LAT, Rsk, CREB
T-cell receptor signaling [16]	CD28, CD4, TCRlig	NFKβ	46	83	62	3	TCRb, ZAP70, DAG, SHP1, SLP76, Lckp1, PAG, Csk, CaM, cCblp2, Vav1, Gab2, Abl, Fyn, ERK, PIP3, PDK1, PLCga, TCRp, LAT, Ikkab, IkB
TGFβ signaling in cell fate change [18]	TGFβ	EMT	19	71	$>2^{16}$	3	GSK3β, GLI, AKT
T lymphocyte differentiation [17]	IL27, IL23	IL17	50	112	$>2^{16}$	4	IL27R, IL23R, STAT1, Tbet, GATA3, RORgt, Foxp3, STAT3
T lymphocyte differentiation [17]	IL27	IL10	47	109	$>2^{16}$	2	IL27R, STAT1, Tbet, GATA3, RORgt, Foxp3, STAT3

Table 4-1. The maximum number of node-independent ESMs obtained for five network models of different biological systems. The name of the network is given in the first column. The second and the third columns specify the input(s) and output for which the functional redundancy of the signaling is calculated. If a combination of multiple inputs is used, all inputs appear in all ESMs identified. The fourth and fifth columns indicate the number of nodes and edges in the sub-network specific to the input(s)/output pair. That is, the columns indicate the union of all nodes and edges in the ESMs identified for the input(s)/output pair. The last column gives the mandatory reusable node

set for each case of node-independent ESM identification. The nodes in this set are located in the signaling cascade between the specific input(s)/output pair, and they appear in all ESMs found. In order for the network to potentially have more than one node-independent ESM, these nodes have to be made reusable for all ESMs.

From Table 4-1 one can observe that some degree of redundancy exists for all networks, as the number of nodes shared by all ESMs is less (in some cases much less) than the number of nodes between the input and the output node. A large mandatory reusable node set in a network reflects the fact that the nodes in that network are strongly functionally dependent on each other. The more Boolean *And* functions there are, the more components are required for successful signal transduction. This is the case for the host-immune response network and the T cell receptor signaling network, whose mandatory reusable node sets include 87.5% and 40.5%, respectively, of the number of nodes between the input and output. It is observable that the maximum number of node-independent ESMs does not grow linearly with the number of ESMs in the system. For network models that have multiple inputs and/or outputs, such as the T cell receptor signaling network and the T lymphocyte differentiation regulation network, the signaling between different input/output pairs may have different degrees of redundancy.

To better illustrate one of these five networks, the Boolean model of T-lymphocyte differentiation [17], on Figure 4-6 we present the connected component of its ESM contingency graph. The figure illustrates the possibility of multiple full sets of node-independent ESMs of different sizes. An ESM (e.g. ESM_1) can be placed with another ESM (e.g. ESM_8) and form a full set of node-independent ESMs, or it can be placed with a different combination of ESMs (e.g. ESM_2 and ESM_7) and form another full set. Being a full set of node-independent ESMs does not guarantee that it is of maximum size. In Figure 4-6, the set $\{ESM_1, ESM_2, ESM_3, ESM_4\}$ is the *maximum* full set.

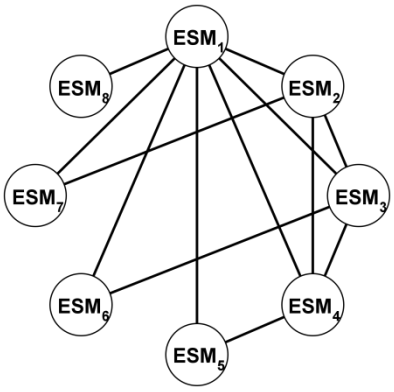


Figure 4-6. The connected component of the ESM contingency graph of the ESMs found for the input combination IL27, IL23 and output IL17 in the T lymphocyte differentiation network model [17]. The reusable node set is [IL27R, IL23R, STAT1, Tbet, GATA3, RORgt, Foxp3, STAT3]. Full sets that contain a single ESM, meaning that the ESM is not node-independent from any other ESMs, are not shown. These would appear as disconnected single nodes on the ESM contingency graph. ESM_1 and ESM_8 make up a full set of size 2. There are three full sets of size 3: {ESM_1, ESM_2, ESM_7}, {ESM_1, ESM_3, ESM_6} and {ESM_1, ESM_4, ESM_5}. The maximum full set is given by {ESM_1, ESM_2, ESM_3, ESM_4}. Any subset of a full set is node-independent as well.

We selected two signal transduction networks, one describing the signaling process of light-induced stomatal opening [19] and the signaling process of TBFβ in hepatocellular carcinoma epithelial-to-mesenchymal transition [18], for a comprehensive analysis.

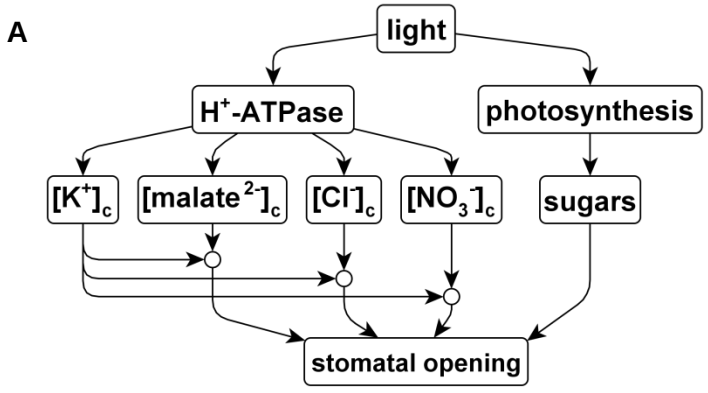
4.2.4.1 The light-induced stomatal opening network

Our starting point is a multi-level discrete dynamic model constructed for the signaling process of light-induced stomatal opening based on an extensive curation of the experimental literature [19]. Light induces stomatal opening via two main pathways: the activation of the proton pump H^+-ATPase which in turn elicits the accumulation of ionic osmotica (e.g., K^+), and the activation of photosynthesis which in turn promotes the accumulation of sugar, a non-ionic osmoticum. In order for charge balance potassium ion requires the presence of one or more anions. Each of the three representative anions is therefore connected with potassium ion with a Boolean *And* function, indicated by the composite nodes (small circles) in Figure 4-7A, B. The

accumulation of osmotica will cause stomatal opening due to the increase in osmotic potential and the uptake of water that ensues. We simplify the original network by focusing on the osmotica directly contributing to stomatal opening and the primary driving force for their respective accumulation, leading to the simplified network shown in Figure 4-7A. In this network the rule for all nodes with a single upstream activator is to follow the value of its activator, i.e. $Y^*=X$, and the Boolean rule for the output node is:

stomatal opening*=$[K^+]_c$ *And* $[malate^{2-}]_c$ *Or* $[K^+]_c$ *And* $[Cl^-]_c$ *Or* $[K^+]_c$ *And* $[NO_3^-]_c$ *Or* sugars.

This simplification process greatly reduces the state space of the model yet the biological fact that stomatal opening can be elicited by the accumulation of either sugars or ionic osmotica is captured, making the model suitable for the application of the node-independent ESM analysis. Simplification such as this may be used in future work aimed to extend the methodology to beyond Boolean dynamic models (see Discussion).



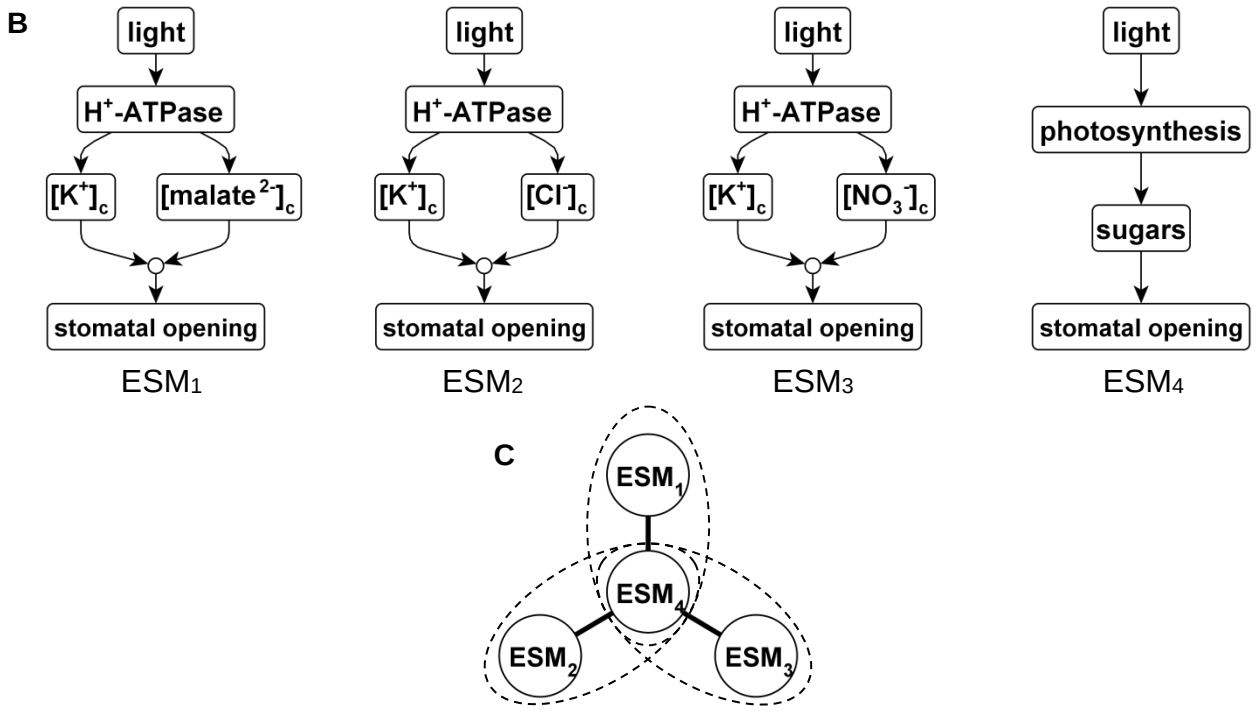


Figure 4-7. A simplified representation of the light-induced stomatal opening network model. (A) The expanded network. Composite nodes (small circles) represent Boolean *And* relations. (B) The four ESMs of the network. (C) The ESM contingency graph. The full node-independent ESM sets are {ESM_1, ESM_4}, {ESM_2, ESM_4}, and {ESM_3, ESM_4}.

Four ESMs have been identified for the simplified light-induced stomatal opening network (Figure 4-7B). Since ESM 1 through 3 share the H^+-ATPase, the primary driving force for ionic uptake, as well as $[K^+]_c$, the predominant positive ion being accumulated in guard cells, they are not node-independent. All three of them, however, are node-independent from the sugar accumulation pathway, characterized by ESM_4. The ESM contingency graph henceforth shows up as a tri-star. This is a special case of the example shown in Figure 4-4 with M=3, N=1 (or vice versa). The three full node-independent ESM sets, {ESM_1, ESM_4}, {ESM_2, ESM_4}, and {ESM_3, ESM_4}, are encircled with dashed lines on Figure 4-7C. The fact that the maximum number of node-independent ESMs among the solutions is two captures a critical aspect of the biology: the accumulation of sugar has a distinctively different mechanism from the accumulation of ions while all ion accumulation shares a common primary driving force. It also implies that in order to

completely shut down stomatal opening, one would have to interfere with not only the H^+-ATPase, which only affects the ionic pathways, but the sugar accumulation pathway as well. Since the two mechanisms are node-independent, the network has a functional redundancy of two, which makes it failure proof to any single node knockout in the network.

4.2.4.2 The TGFβ signaling in hepatocellular carcinoma epithelial-to-mesenchymal transition network

Steinway et al. [18] constructed a Boolean model of TGFβ signaling during hepatocellular carcinoma epithelial-to-mesenchymal transition (EMT). This transition is a cell fate change from a cuboidal, less mobile and therefore localized cell type into an irregularly-shaped, more mobile and therefore non-localized cell type. This cell fate change mainly happens in embryonic development and wound healing, but it can happen aberrantly in cancer, allowing cancer cells to leave the primary tumor site and establish distant metastases. This process can be triggered by a number of known signals, among which TGFβ is a main signal of interest. The model recapitulates known signaling events and dysregulations during the induction of EMT. It also predicts the activation of the Wnt and Sonic Hedgehog signaling pathways during this process. The starting point of our analysis is a simplified and TGFβ-focused version of the network, constructed by Steinway et al., which was shown to capture the TGFβ-induced dynamics of the full network [18]. The expanded network of the TGFβ signaling specific network has 19 real nodes, 19 composite nodes and 105 edges (Figure 4-8). More than 65 thousand ESMs have been identified for the expanded network.

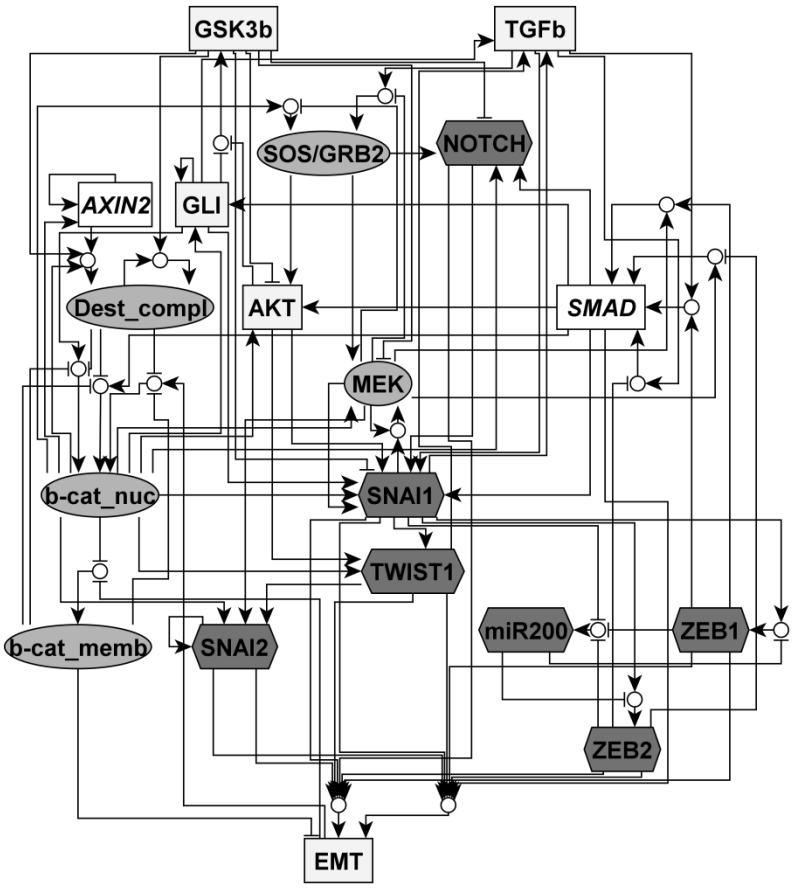


Figure 4-8. A full set of node-independent ESMs in the network model of TGFβ signaling in EMT. Edges that terminate in arrows represent activation; edges with terminating segments represent inhibition. The composite nodes are represented with small circles. There are four types of node symbols: plain rectangles with light background represent the input and the output nodes of the network (TGFβ and EMT) and the mandatory reusable node set (GSK3β, GLI and AKT). Oval shaped nodes with grey background are specific to ESM_1, and hexagonal nodes with dark grey background are specific to ESM_2. Rectangles with white background and italic font (AXIN2, SMAD) are not included in either ESM in this solution set of node-independent ESMs.

Apart from the input, TGFβ, and the output, EMT, of the network, three nodes appear in all ESMs, and they are put into the reusable node set accordingly: GSK3β, GLI and AKT. The fact that these three nodes appear in all ESMs signifies their importance for TGFβ signal transduction. Using only these three nodes in the reusable set, we find that the maximum number of node-independent ESMs is two. Figure 4-8 illustrates one of the possible full sets. Oval shaped nodes are specific to ESM_1, and hexagonal nodes are specific to ESM_2. Nodes with white background

and italic text are not included in either ESM_1 or ESM_2 (they are, however, included in other ESMs).

An interesting observation in the TGFβ network is that the node SMAD is not part of the mandatory reusable set. Indeed, it is not included in either ESM shown in Figure 4-8. The SMAD complex is a canonical mediator of TGFβ signaling [18] and many would expect that its disruption eliminates TGFβ signaling. However, our analysis indicates, without doing any dynamic simulations, that disrupting the node SMAD will not eliminate all ESMs (e.g. the two ESMs shown on Figure 4-8 will be preserved) and thus will not drastically disrupt TGFβ signal transduction. This result is corroborated by the dynamic simulations of Steinway et al. [18], predicting that SMAD knockout only leads to a minor reduction of the probability of the cell fate change. Moreover, recent experiments indicate that downregulation of *smad* transcription by siRNA indeed leads to a minor reduction in the cell fate change in a liver cancer cell line, validating the theoretical prediction.

4.2.5 Node-independent ESMs quantify node importance

The ESM is a powerful concept that is able to characterize the functional connectivity of an input-output pair. Due to the overlaps between ESMs, disruption of a node may disconnect multiple signal carrying ESMs of the system. Indeed, the relative reduction in the number of ESMs due to the disruption of a node was shown to be a successful measure of the node's functional importance [6]. The identification of the maximum set(s) of node-independent ESM of a system allows a strong insight into the expected effect of a node disruption. If a node that is exclusive to one of the ESMs in this set is disrupted, it is expected to have no or little effect on the signal transduction since other independent ESMs exist. If, on the other hand, a node in the (non-empty) mandatory reusable set is disrupted, signal transduction is completely disrupted. For a validation

of this statement, we compare our results to the previous results of ESM-based node importance in [6]. For the three network examples examined in [6] (the host immune response network [15], the guard cell ABA signaling network [14] and the T-cell receptor signaling network [20]), the mandatory reusable node set identified for each of the network is identical with the set of nodes assigned the highest importance, 1, in [6].

Therefore, if a network has an empty mandatory reusable node set, the signaling process represented by it is robust against any single node disruptions. By applying our method to established network models of biological signaling processes, we found that the mandatory reusable set is usually non-empty.

4.3 Discussion

In this study, we propose a measure of the redundancy of signal transduction networks. This measure builds on a network representation that reflects not just who interacts with whom but also the combinatorial effects of multiple regulators. It also uses a measure of input-output connectivity, the elementary signaling mode, better suited to signal transduction networks than simple paths. We find that most biological signal transduction networks are not functionally redundant, as they don't have more than one node-independent elementary signaling mode. We deepen the analysis of these networks by introducing the new concept of mandatory reusable node set. This allows the identification of nodes that are of the highest importance to the signaling process without carrying out any dynamic simulations. The size of the mandatory reusable node set is closely related to the functional dependencies between pathways. If all nodes in a network are functionally dependent on each other, i.e. all nodes are connected with Boolean *And* or *And Not* functions, the maximum number of node-independent ESMs is one and the network has no functional redundancy. In order for a network to have two or more node-independent ESMs, the

network is required to contain at least one Boolean *Or* function, and the largest possible size of the mandatory reusable node set equals the number of nodes between the input and the output minus one.

The agreement of our analysis with much more effort-intensive methods, both on the criticality of certain nodes and on the non-criticality of others, demonstrates the predictive power of the concepts and measures introduced here.

We note that using the mandatory reusable node set does not ensure that a network has a maximum full set of more than one node-independent ESM. In other words, allowing nodes that appear across all ESMs to be reused does not guarantee that the system will have more than one node-independent ESM. A related example is given in Figure 4-9, which illustrates a network with an empty mandatory reusable set. Since the mandatory reusable node set is empty, one can predict that the signal transduction is robust against any single node knockouts (see Figure 4-9B). However, none of the three ESMs of Figure 4-9 are node-independent of the other two.

We also note that as more nodes are being added to the reusable node set, the number of ESMs in the maximum node-independent ESM set is not guaranteed to increase. For example in Figure 4-9 adding node A, then B, then C to the reusable node set does not lead to an increase in the number of node-independent ESMs. It can be concluded from this example that entire paths rather than individual nodes have to be added to the reusable node set before the maximum number of node-independent ESMs increases. The structure formed by the reusable nodes plays a more important role than their number.

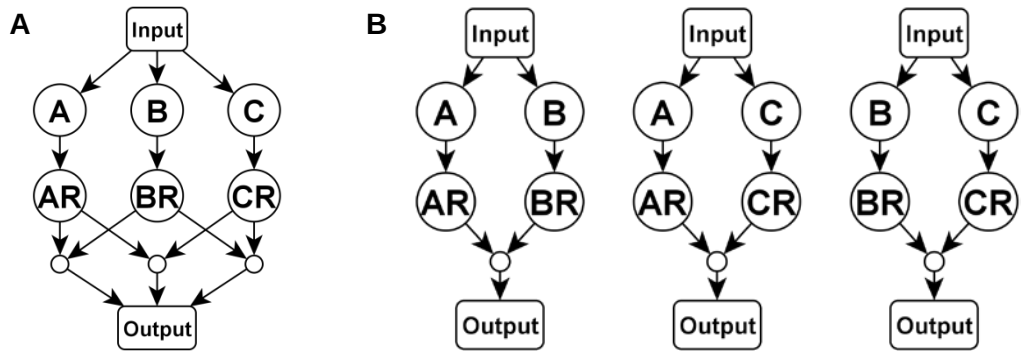


Figure 4-9. An example illustrating the relation between the size of the reusable node set and the maximum number of node-independent ESMs of a network. The example network shown in (A) consists of 3 nodes: A, B, C, with their respective response elements, AR, BR, CR. The composite nodes signify the Boolean *And* relations. The Boolean transfer function for the output node is: Output*=(AR *And* BR) *Or* (AR *And* CR) *Or* (BR *And* CR). The ESMs of the network are shown in (B). The mandatory reusable node set is empty. The maximum number of node-independent ESMs is one. If we successively add node A, B, and C to the reusable node set, the three ESMs of the network continue to be node-dependent from each other, since they still share node AR, BR, or CR.

An interesting question related to the structure of the ESM contingency graphs remains open: is an arbitrary graph a viable structure as an ESM contingency graph? For instance, we have seen a cycle of 3 nodes (N=3, Figure 4-4) and a cycle of 4 nodes (M=N=2, Figure 4-5) as an ESM contingency graph. The original networks behind these two ESM contingency graphs, however, bear little resemblance to each other. Is a cycle of 5 nodes, or generally, a cycle of N nodes achievable as an ESM contingency graph?

The measure we proposed in this work is applicable to network models with Boolean dynamics in its current form, as the definition of the expanded network and of the ESMs requires Boolean operators. Discrete dynamic models can be transformed into Boolean models by representing multi-state nodes with a group of binary nodes [21]. Continuous variables can be binarized by using threshold values based on domain knowledge, and thus continuous network models can be transformed into Boolean models. The functional redundancy of the derived

Boolean network reflected by its node-independent ESMs can subsequently be interpreted in the context of the original model. Alternatively, a generalization of the OR relationship (independent regulators, each sufficient) and the AND relationship (all regulators need to act together) to other frameworks can lead to a generalization of the ESM concept and to a straightforward application of our methods. Another area of future development is therefore to extend the measure to models with multi-state or even continuous dynamics.

4.4 References

1. Klamt S, Saez-Rodriguez J, Lindquist JA, Simeoni L, Gilles ED (2006) A methodology for the structural and functional analysis of signaling and regulatory networks. BMC Bioinformatics 7: 56.
2. Klamt S, Saez-Rodriguez J, Gilles ED (2007) Structural and functional analysis of cellular networks with CellNetAnalyzer. BMC Syst Biol 1: 2.
3. Goemann B, Wingender E, Potapov AP (2009) An approach to evaluate the topological significance of motifs and other patterns in regulatory networks. BMC Syst Biol 3: 53.
4. Wunderlich Z, Mirny LA (2006) Using the topology of metabolic networks to predict viability of mutant strains. Biophys J 91: 2304-2311.
5. Ma'ayan A, Jenkins SL, Neves S, Hasseldine A, Grace E, et al. (2005) Formation of regulatory patterns during signal propagation in a Mammalian cellular network. Science 309: 1078-1083.
6. Wang RS, Albert R (2011) Elementary signaling modes predict the essentiality of signal transduction network components. BMC systems biology 5: 44.
7. Zanudo JG, Albert R (2013) An effective network reduction approach to find the dynamical repertoire of discrete dynamic networks. Chaos 23: 025111.
8. Albert R, Othmer HG (2003) The topology of the regulatory interactions predicts the expression pattern of the segment polarity genes in Drosophila melanogaster. Journal of theoretical biology 223: 1-18.
9. Samaga R, Saez-Rodriguez J, Alexopoulos LG, Sorger PK, Klamt S (2009) The logic of EGFR/ErbB signaling: theoretical properties and analysis of high-throughput data. PLoS computational biology 5: e1000438.
10. Albert R, Thakar J (2014) Boolean modeling: a logic-based dynamic approach for understanding signaling and regulatory networks and for making useful predictions. Wiley interdisciplinary reviews Systems biology and medicine 6: 353-369.
11. Wang R, Sun Z, Albert R (2013) Minimal functional routes in directed graphs with dependent edges. International Transactions in Operational Research 20: 19.
12. Milo R, Shen-Orr S, Itzkovitz S, Kashtan N, Chklovskii D, et al. (2002) Network motifs: simple building blocks of complex networks. Science 298: 824-827.
13. Lipshtat A, Purushothaman SP, Iyengar R, Ma'ayan A (2008) Functions of bifans in context of multiple regulatory motifs in signaling networks. Biophys J 94: 14.
14. Li S, Assmann SM, Albert R (2006) Predicting essential components of signal transduction networks: a dynamic model of guard cell abscisic acid signaling. PLoS biology 4: e312.
15. Thakar J, Pilione M, Kirimanjeswara G, Harvill ET, Albert R (2007) Modeling systems-level regulation of host immune responses. PLoS computational biology 3: e109.

16. Saez-Rodriguez J, Simeoni L, Lindquist JA, Hemenway R, Bommhardt U, et al. (2007) A logical model provides insights into T cell receptor signaling. PLoS computational biology 3: e163.
17. Martínez-Sosa P, Mendoza L (2013) The regulatory network that controls the differentiation of T lymphocytes. Bio Systems: 8.
18. Steinway SN, Zanudo JG, Ding W, Rountree CB, Feith DJ, et al. (2014) Network Modeling of TGFbeta Signaling in Hepatocellular Carcinoma Epithelial-to-Mesenchymal Transition Reveals Joint Sonic Hedgehog and Wnt Pathway Activation. Cancer research 74: 5963-5977.
19. Sun Z, Jin X, Albert R, Assmann SM (2014) Multi-level Modeling of Light-Induced Stomatal Opening Offers New Insights into Its Regulation by Drought. PLoS computational biology 10: e1003930.
20. Zhang R, Shah MV, Yang J, Nyland SB, Liu X, et al. (2008) Network model of survival signaling in large granular lymphocyte leukemia. Proceedings of the National Academy of Sciences of the United States of America 105: 16308-16313.
21. Miskov-Zivanov N, Turner MS, Kane LP, Morel PA, Faeder JR (2013) The duration of T cell stimulation is a critical determinant of cell fate and plasticity. Science signaling 6: ra97.

Chapter 5

Determining the Attractors of a Boolean Network Using an Elementary Signaling Mode Approach

5.1 Introduction

It is increasingly accepted that the analysis of complex systems encountered in many fields of science and engineering is greatly simplified by representing the system as a network. The elements of the system become network nodes and the relationships among the elements become the edges of the network. Examples of such systems fruitfully analyzed as networks include biological systems at several levels of organization (1, 2), various social systems (3), the World-Wide Web (30), as well as massive engineered systems (31). Network analysis and modeling has played an important role in many fields. In biology, for instance, network theory offers tools to describe the mass flow or regulatory relationships of diverse molecules and cellular components and to reveal complex cellular functionalities that emerge from these molecular interactions (4-7, 35).

A network is a set of nodes (or vertices) that represent the elements in a system and a set of edges (or links) that stand for the relationships among nodes. The edges can be undirected, indicating the adjacency relationship between nodes, or directed, representing causality, ordering or regulation in a system. Further, certain directed edges may also be distinguished by signs, if a node can act as either an activator or an inhibitor of another. Our focus here is on directed and signed networks, which are encountered in biological systems (e.g. signal transduction networks, the immune response to pathogens, ecological systems that involve both predator-prey and mutualistic interactions) and in social systems (e.g. inter-personal networks that include both

friendship and enmity). Most directed networks encountered in applications contain a number of source nodes (nodes with no incoming edges), which can be considered as entry points (inputs) to the system, and a number of sink nodes (nodes with no outgoing edges), which can serve as outputs of the network.

Many complex systems serve as substrates for spreading or flow processes. Examples include disease or rumor spreading, cascading failures, or information spreading on biological signal transduction networks. In these cases in addition to the network structure, which captures the topological layout of a system, the capability to elucidate the temporal behavior of the system is often desirable, leading to the construction of a layer of dynamic models on top of the static network structures. Dynamic models characterize the nodes by a state variable (for example, categorizing individuals into susceptible, infected or recovered in disease spreading models, or describing molecules by their concentrations in chemical reaction networks). The regulatory effects of edges incident on a given target node are synthesized into a so-called transfer function that indicates the relationship between the target node's state and its regulators' states.

Discrete dynamic models are increasingly emerging as practical and successful simplified representations for the dynamics of various systems (1, 2, 9-14). Boolean dynamic modeling is the simplest type of discrete dynamic model, in which the node states are represented as binary values, namely 1 (indicating an above threshold level or activity) and 0 (below-threshold level or inactivity). A Boolean transfer function (or Boolean rule) is expressed using the logical operators "not", "and", "or" or as a threshold rule (11, 36, 37). Dynamic simulations performed based upon Boolean models yield the long-term behaviors of the system, also known as attractors. For a system of N nodes, the total size of its *state space* is 2^N. If the simulation is initiated from a state (called *initial state*), the system will likely visit a series of other states, leaving a *trajectory* in the state

space. An attractor will be reached eventually. Such a single run is unlikely to cover the *entire* state space especially when N is large. In many systems there isn't enough information on the most relevant initial state, thus a systematic exploration of the state space is needed, which is time-consuming or even improbable for large N's.

It is now widely recognized that the structure of a network is closely related to its function (15-18). For example, particular interaction patterns, called network *motifs* (19), are over-represented in various systems, and they have been proposed to achieve certain functions and generate characteristic attractors (20). Here we aim to determine the attractors of a Boolean network based purely on its network topology and the Boolean transfer functions of the nodes. For the majority of biological networks the most relevant attractors are steady states that correspond to robust responses to sustained signals. Therefore, we are focusing on the case wherein at least one the input node is in the state 1 (there is a sustained signal), and we are most interested in the steady states obtained in this case. We pursue other types of attractors of the system only in cases where steady states do not exist. The method in general, however, is able to retrieve all attractors of a system independent of its signals. Our method has three principle underpinnings: the concept of elementary signaling modes (ESM) proposed in (21), the close relationship between the degree of sign-consistency of a network and its attractors, and the critical role of the positioning of the strongly connected components (SCC) in a network.

Our method fills a blank in Boolean dynamic modeling, not only producing an alternative approach to solve for the attractors, but also offering new insights into the relationship between network topology and its dynamics, generating a rich repertoire of results concerning sign-consistencies of an ESM, potentially feasible steady states of a sign-consistent ESM and network simplification.

5.2 Background and methods

5.2.1 The benefits of Boolean dynamic modeling

Many traditional dynamic models are continuous and deterministic; they describe node states as continuous variables (e.g. abundances or concentrations) and take the form of a set of coupled differential equations. Each differential equation describes the rate of change of a regulated node's state as a function of the states of its regulators (8). Approaches of such fine granularity of time and node state, however, grow exponentially taxing on computing power and become practically not feasible to carry out when the number of nodes in a system reaches over a hundred. Even for smaller systems, the estimation of kinetic parameters is often difficult due to insufficient temporal data or unknown mechanistic details. Alternative dynamic approaches that overcome these difficulties are based on the qualitative states observed in many different contexts. For example, transcription factors may be post-translationally activated or suppressed, or an ion channel is either open or closed. In other situations, though the node states are continuous (e.g. concentrations), their exact value is of less interest than whether they are above a certain threshold. These observations give rise to one type of discrete dynamic modeling—Boolean modeling (9, 10), which has been successfully implemented on a wide range of biological systems, capturing known characteristics and features, offering insights as well as yielding novel predictions (1, 2, 11-14).

5.2.2 State trajectories and attractors in Boolean models

Assuming that the total number of nodes in a system is N, a *state of the system* is an N-dimensional vector, with each component equal to the state of a node, which in turn is equal to 1 or 0. During dynamic simulations of the system, its state will evolve following the set of Boolean transfer functions that govern the state transitions of each node. An *attractor* is a steady

(unchanged) state or a set of repeating states of the system. For the latter case, called a *complex attractor*, every state in the set will eventually reappear in the state trajectory as the system evolves dynamically through time. Both types of attractor represent stabilized system behavior, and can be mapped onto and thus confirmed by experimental results. The steady states of a molecular network could correspond to cell phenotypes (13, 32). Complex attractors can represent circadian rhythms (33, 34), cell cycles (11, 12), or seasonal variations in ecological communities. The fact that the attractors of a Boolean dynamic model can be interpreted experimentally underscores the importance of determining all possible attractors of a Boolean model.

5.2.3 Strongly connected components and sign-consistency

In a directed graph, a strongly connected component (SCC) is a group of nodes within which a path starting from any node i and ending at any node j exists. i and j can be identical, meaning the path can start and end at the same node. In other words, any node within an SCC can be reached via a directed path from itself or any other node in the same SCC. The in-component of an SCC is defined as the set of nodes that can reach the SCC (that is, for which at least one path exists that starts at the node and ends at a node of the SCC). The out-component of an SCC is the set of nodes that can be reached from the SCC.

A directed and signed graph G is sign consistent if (and only if) for all pairs of nodes i and j in G, all paths that exist from i to j have the same sign, and all cycles in G are positive (22). The sign of a path can be obtained by counting the number of negative edges on that path: if the number of negative edges on the path is odd, the path is negative; if the number of negative edges on the path is even, the path is positive. The same definition applies for cycles, except that for a cycle the counting has to start and end on the same node in the cycle. Sign-consistency means an

unambiguous positive or negative relationship between any pair of nodes.Biological networks are close to being sign-consistent; for example, the percentage of the sign-consistent sub-networks of transcriptional regulatory networks mapped to date ranges between 80-90% (28).

The concept of sign-consistency has been fruitfully used in many fields (22, 23). The theory of both continuous and discrete dynamical systems indicates that a necessary condition for complex dynamic attractors is the existence of sign-inconsistent feed-forward or feed-back loops in the network (20, 27). Sign-consistent networks behave in many ways like one-dimensional systems, for example do not exhibit stable oscillatory behaviors (23). Correspondingly, a sign-consistent Boolean network can only have steady states as its attractors, whereas a sign-inconsistent Boolean network can also have complex attractors.

The essence of sign-inconsistent networks can be condensed into two simple network motifs (Figure 5-1).

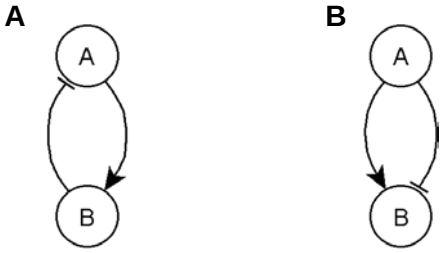


Figure 5-1. Two network motifs encountered in sign-inconsistent networks. (A) A negative feedback loop, which leads to complex attractors (B) A sign-inconsistent feedforward loop which leads to an 'isolation' type steady state.

In the shortest possible negative feedback loop, illustrated in Figure 5-1A, the state transition of node A is governed by A* = *not* B, where the asterisk refers to the next state of A, and the next state of B is B* = A. This set of transfer functions does not have a steady state, since the set of equations that results when taking time away yields B = *not* B, which does not have a

solution. Therefore, the system can only have a complex attractor. The sign-inconsistent feedforward loop, in which there are two independent paths of opposite sign between a source node and a target node, will also generate undesired network behavior (29). In the shortest possible sign-inconsistent feed-forward loop, illustrated on Figure 5-1B, for instance, the Boolean transfer function for node B is B* = A *and not* A, which can be in turn simplified into B* = 0. The converse case of B* = A *or not* A yields B*=1. In both cases the steady state of the target node is independent from the state of the source node. For this reason we denote the steady state resulting from a sign-inconsistent feed-forward loop an isolation type steady state.

A sign-inconsistent graph can be modified into a sign-consistent one by removing a set of edges, or changing the signs of a set of edges. We use the implemented algorithm of DasGupta, B., *et al* (22) to test the sign-consistency of networks and the minimum set of edges whose removal will make the network sign-consistent Note that when a sign-inconsistent edge is removed, only the relevant 'relation' between the corresponding nodes is disregarded. The starting and ending nodes of the edge are kept in the network, maintaining their remaining edges.

5.2.4 Elementary signaling modes

The most frequently used graph measures, such as degree distributions, path lengths, centrality measures, do not distinguish between edge signs and implicitly assume that each edge is independent of the others. To account for the special properties of signal transduction networks, in (21) an expanded network representation was proposed. In this representation complementary nodes were introduced for components that play a role in negative regulation and composite nodes were introduced to express synergy in a graphical form. Specifically, given a Boolean transfer function for a target node written in a disjunctive normal form (DNF) (Table 5-1), a new composite

node will be created for each clause in the rule. This composite node will have as incoming edges the set of edges corresponding to the respective clause, and will have a positive outgoing edge directed to the original target node. Here we follow (21) in augmenting the network with composite nodes, but do not introduce complementary nodes and keep the edge signs. Original nodes that were present in the original network together with composite nodes form the expanded network.

A Boolean transfer function, not in DNF	D = (A *or not* B) *and* C
The same Boolean transfer function, in DNF	D = (A *and* C) *or* (*not* B *and* C)
The two clauses in the Boolean transfer function	(A *and* C), (*not* B *and* C)

Table 5-1. In Boolean logic, a disjunctive normal form (DNF) is the writing of a Boolean transfer function (Boolean equation) as a set of Boolean clauses, each clause containing only the *not* and/or *and* operators, connected with the *or* operator.

An elementary signaling mode (ESM) was defined in (21) as the minimal set of components in a signal transduction network that allow information propagation from the source node (signal) to the output node. The ESM is minimal in a similar sense as an elementary flux mode in metabolic networks (17): it is not decomposable and knockout of any of the nodes in the ESM will make it unable to transduce the signal. The ESM concept generalizes the concept of shortest path; the difference between the two concepts is that an ESM that contains a composite node needs to also contain all nodes directly upstream of it. Multiple edges that converge on the same target node and that are inside the same ESM are in a logical AND relation (because they are converging on a composite node), and separate ESMs affect the output node in an OR relation.

5.3 Results

Our procedure for determining the attractors of a signal transduction network has the following main steps:

Network preprocessing

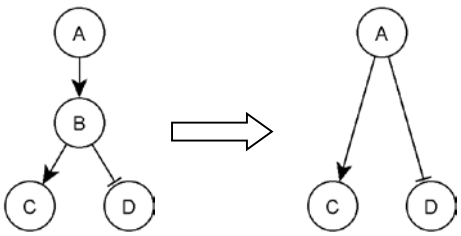
Listing all ESMs

Determining the attractor(s) of Network

5.3.1 Network preprocessing

For a Boolean signal transduction network of N nodes, the dimension of a single state vector of the system is N. Since every node of the network will be involved in any attractor identification procedure, it is very beneficial to decrease the complexity of the problem by simplifying the network without reducing its attractor repertoire. This will also greatly expedite the calculation, save space in terms of memory and henceforth enhance the performance of any algorithm implemented. A previously introduced network reduction strategy is to merge a node with a single, positive, incoming edge with its sole upstream regulator (24, 25). As this merging only involves simple substitution, it preserves the number of attractors of the system(26, 38, 39). Adapting this strategy, we simplify the network by repeatedly: i) remove a node that has a single upstream activation edge, and connect all its downstream targets to its sole upstream node; ii) remove duplicate edges of identical sign between a pair of nodes (Figure 5-2).

A



B

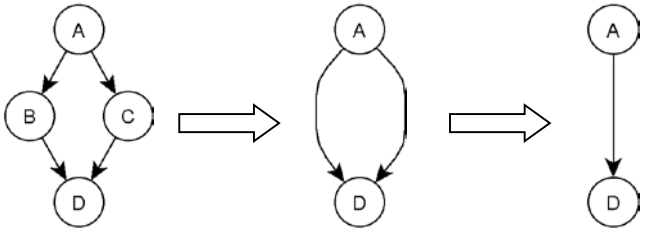


Figure 5-2. Network simplification while the total number of attractors of the system is preserved. In a directed and signed graph, a node can act either positively (activation, corresponding edge usually denoted with "→") or negatively (inhibition, corresponding edge usually denoted with "—|") on another node. (A) Node B has a sole, positive, regulator, node A, thus it can be eliminated, or equivalently, merged with node A. The elimination of B rewires all its original downstream edges, to nodes C and D, to its upstream node A. (B) If duplicate edges emerge after node elimination, they are merged into a single edge.

After the network is simplified, it is expanded to represent the synergistic relationships (AND rules) between nodes by the addition of composite nodes (21, see Background and Methods).

5.3.2 Identifying all ESM's

Given the expanded Boolean network, we use the bottom-up (that is, backwards from the output node) ESM growing algorithm proposed in (40) to identify all existing ESMs of the network. The method is succinctly described as follows: the algorithm examines the nature of the node most recently added to the ESM (current node). If the node is an original node, indicating that its *n* upstream nodes are joined by a Boolean *or* function and can henceforth transduce information independently from each other, *n* partial ESMs will be generated with a different upstream node attached to the current node in each partial ESM. If the node is a composite node, indicating that its upstream nodes are joined by Boolean *and* functions, all upstream nodes are required to transduce information, thus a single partial ESM will be generated with all upstream nodes

attached to the current node. The method is iteratively carried out to nodes that are newly attached in a partial ESM until the partial ESM stops growing in size and becomes completed.

5.3.3 Solving for the attractors of a Network

In order to solve for the attractors of a network, several sub-steps are taken:

Calculate the sign-consistency of the expanded network and determine the minimum set of sign-inconsistent edges whose removal renders the rest of the graph sign-consistent;

Depending on the number of sign-inconsistent edges removed, the following measures are taken:

i) 0 sign-inconsistent edges. The expanded network is sign-consistent. The attractors are obtained by an algorithm based on the network's SCC layout.

ii) 1 sign-inconsistent edge. The expanded network is sign-inconsistent. The sign-inconsistent edge is temporarily removed from the network, leaving the rest sign-consistent. The attractors of the sign-consistent part of the network are obtained using the same method as used in case i). The sign-inconsistent edge is then put back into the network, and the attractors for the whole network are calculated.

iii) 2 or more sign-inconsistent edges. The network is sign-inconsistent. The attractors of the network are solved using a state backtracking—consistency check algorithm we developed.

We will look at the implementation and the argument behind each step in detail.

5.3.3.1 Determining the attractors of a sign-consistent Boolean network

In sign-consistent networks, all paths between a pair of nodes have the same sign and all cycles are positive, these networks are composed of ‘islands’ that consist of maximal sets of connected positive edges, bridged by negative edges if there are any (Figure 5-3). An island can be a single node, a single SCC, multiple connected SCCs with in-components and/or out-components, or a tree-like structure without any SCC present. If there are two or more islands, there is no requirement on the number or direction of the negative bridges: there doesn’t have to be a bridge between every pair of islands (although they have to be connected through other islands), there can be a single bridge directed from one island to another, or multiple bridges in either or both directions. The only requirement is that there cannot be a cycle formed by an odd number of islands connected by negative edges, since this cycle would be negative.

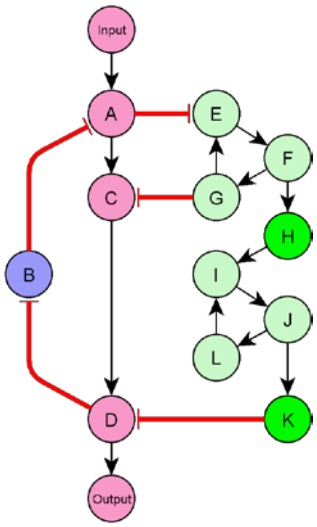


Figure 5-3. Illustration of a sign-consistent network. Three islands that contain only positive edges exist, marked with blue, pink and green, respectively. The blue island is composed of a single node; the pink island is a simple tree-like structure while the green island contains two SCCs, EFG and IJL, marked with a lighter green, with two connection nodes marked with darker green. Negative edges (red-colored) form the ‘bridges’ between the islands. Bridges exist between the blue and pink, pink and green islands. The blue and green islands are not directly

connected by bridges. All paths between any given pair of nodes in the network are sign-consistent with each other and all existing cycles in the network are positive, thus the whole network is sign-consistent.

From the structural analysis above it can be seen that the elements most crucial to determining the attractors of a sign-consistent ESM is the existence of SCCs in a network, the relative positioning among all SCCs, and the alignment of negative edges with SCCs. We hereby introduce two lemmas concerning the attractor properties of SCCs.. Recall that inside an ESM, multiple incoming edges to the same node are connected by the Boolean *AND* function.

Lemma 1: An SCC of all positive edges inside an ESM can have two steady states, namely all node states = 1 and all node states = 0, and has no other attractors.

Lemma 2: The all node states = 1 steady state of the SCC of all positive edges within an ESM is conditioned on the states of the nodes in the in-component of the SCC. All external inputs to the SCC have to be activations in the all states = 1 steady state. Conversely, the all node states = 0 steady state of the SCC is independent of the states of the nodes in the in-component of the SCC (Figure 5-4).

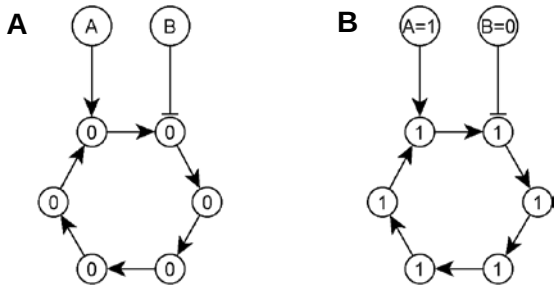


Figure 5-4. An illustration of the steady states of an SCC with positive edges and two external inputs, A and B, all within an ESM. (A) The all nodes states = 0 steady state. Based on the Boolean equality 0 *and* ANY STATE = 0, any state of node A and B is compatible with the

steady state, because every node has at least one 0 input from within the SCC. (B) The all nodes states = 1 steady state. Since 1 *and* 1 = 1, 1 *and* 0 = 0, only A = 1 and B = 0 are compatible with the steady state.

Based on the two lemmas, the steady states of a sign-consistent ESM can be solved following a simple procedure: first, all negative edges in the ESM are temporarily disregarded, henceforth all the islands and SCCs composed of positive edges remain. Assume the number of islands is *m* and the number of SCCs is *n*. The number of SCCs located inside a single island ranges from 0 to *n*. According to lemma 1 each SCC can have two states, all zero or all one. Since the steady states of tree-like structures of nodes are determined by the totality of all their upstream SCCs, a set of 2^n candidate states consisting of a set of all possible combinations of 0s and 1s on *n* slots is generated, each for one of the SCCs. Now reconsider the negative edges. The relation between two connected SCCs, mediated by either a single edge or a tree structure of nodes, can be activation or inhibition, either directed or bidirectional. Given the nature of the interactions between the SCCs, all steady state candidates are then tested against lemma 2: a candidate is considered a solution if it is consistent with lemma 2 and discarded otherwise. If two SCCs are not connected, no steady state requirement is imposed. Once all the consistent states for all SCCs are obtained, the steady states of the rest of the nodes are deduced accordingly.

Once the attractors of the individual ESMs are identified, the steady states of the original network are obtained in a top-down sweep manner. If a node has two different steady states from two ESMs, the result depends on where the node is located: if the node is within an SCC, both steady states are valid, and the number of the steady states of the system doubles; otherwise, the steady state of the node is 1, downstream nodes states are deducted accordingly, and steady state 0 along with all its downstream components are discarded.

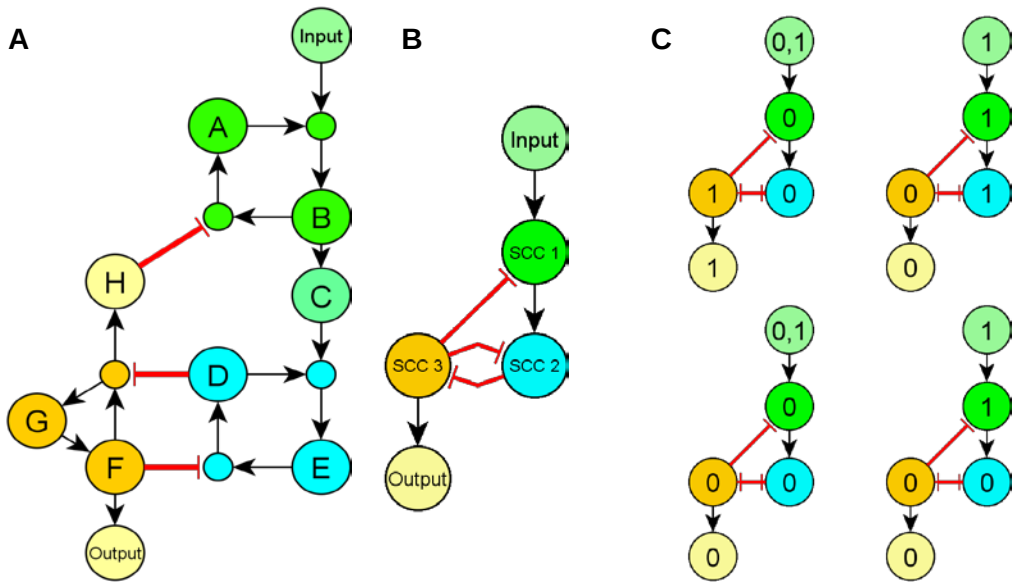


Figure 5-5. Determining the steady states of an example ESM. SCCs consisting of positive edges are marked with different colors, respectively (green, blue and yellow). Nodes that are outside of positive SCCs are marked with a lighter color. (A) The sign-consistent ESM. The composite nodes are where multiple edges join together and the states of their upstream nodes are connected with Boolean *And* function are represented with smaller circles. (B) The relationship among the positive SCCs, the input and output node are. (C) The viable steady states of the ESM based on the two lemmas. The states of the input, output and nodes within the same SCC of positive edges are marked on the corresponding node. Multiple input states indicate that the steady state of the ESM is viable under both input states. The steady states of all nodes in the ESM can be easily retrieved, with node C following the steady state of SCC 1 (green) and node H following the steady state of SCC 3 (yellow).

5.3.3.2 Determining the attractors of an ESM with a single sign-inconsistent edge

If the ESM being solved has a single sign-inconsistent edge, after the steady states for the sign-consistent part of the ESM are determined, it is time to inspect the positioning of the sign-inconsistent edge relative to the existing ESM structure and determine the change it will bring to the steady states. The possible steady states of the source node and target node of the sign-

inconsistent edge have been obtained during the last step. As summarized in Table 5-2, six of the eight steady state combinations of the node pair are compatible with the sign-inconsistent edge.

	Possible combinations of the steady states of the node pair	Compatibility
Activation Edge	0→0	√
	0→1	×
	1→0	√
	1→1	√
Inhibitory Edge	0—|0	√
	0—|1	√
	1—|0	√
	1—|1	×

Table 5-2. Viable steady states of the source and target node of the sign-inconsistent edge.

For instance, if the source node is in state 0 and the target node is in state 1, then putting back a sign-inconsistent negative edge will not make the existing steady state change, or equivalently the steady state found by the last step is still a valid solution for the sign-inconsistent ESM. The viability of two special cases (1→0 and 0—|0) are straightforward based on lemma 2. The target node is in state 0, indicating they are turned off through other sources prior to the addition of the sign-inconsistent edge. Inside an ESM, adding an additional effective activation will not change the 0 steady state of the target node, which also leads to the interesting observation that if the sign-inconsistent edge is ending at a node in state 0, the current state vector is a steady state of the network regardless of the sign of the sign-inconsistent edge.

As has been theoretically and experimentally demonstrated to be true under numerous situations, the presence of a negative feedback loop in a directed and signed network is a necessary condition for oscillatory attractors (20). In particular, if a negative feedback loop is formed by putting a sign-inconsistent edge back into a network and the states of its source and target nodes match one of the two cross-marked cases indicated in Table 5-2, a complex attractor will result. In order to determine the system behavior following the addition of the incompatible sign-inconsistent edge, the original ESM without the sign-inconsistent edge is subject to a new round of SCC determination. This is different from the SCC finding used in step 1 in that all negative edges were disregarded then but included in the finding and grouping this time. After all SCCs are found, all nodes within the same SCC are grouped together and represented by a single 'SCC node' with labeling, thus making the whole ESM a directed acyclic graph (DAG). Based on the DAG, the relative positioning of the source and target nodes will fall into one of the following four categories: i) they are located within the same SCC; ii) they are not within the same SCC, and a path from the source node to the target node exists; iii) they are not within the same SCC, and a path from the target node to the source node exists; iv) they are not within the same SCC and a path in neither direction exists. In case i) and ii), since there is already a path from the source node to the target node, the addition of a sign-inconsistent edge will result in an inconsistent feedforward loop, generating an isolation type of steady state for the target node (see Background and Methods). In case iv), the addition of a sign-inconsistent edge will complete an inconsistent feedforward loop that starts from the input of the network and ends in the target node, which will also give rise to an isolation type of steady state. In case iii), the addition of the sign-inconsistent edge will introduce a negative feedback loop into the network, leading to a complex attractor. In case i), ii) and iv), the steady states of the nodes that are downstream of the isolated node will be updated, and can be

solved accordingly using the new steady state of the isolated node. In case iii), all nodes included in the negative feedback loop will oscillate (switch between states regularly or irregularly). The states of the nodes that are downstream of the oscillating nodes can also be updated following the corresponding Boolean rules, generating a new attractor for the ESM with a single sign-inconsistent edge. Oscillation (Osc) can be deemed as a special 'state' of nodes that is intermediate between 1 and 0, with all Boolean rules involved listed below in Table 5-3. Once the attractors of the individual ESMs are identified, the attractors of the original network can be obtained by a method similar to the one stated for sign-consistent networks, with the exception that the attractor of a node outside any SCC will now follow the rules listed in Table 5-3.

Boolean Rules	*AND*	*OR*	*Not*
All combinations	Osc *AND* 0 = 0	Osc *OR* 0 = Osc	*Not* Osc = Osc
	Osc *AND* Osc = Osc	Osc *OR* Osc = Osc	
	Osc *AND* 1 = Osc	Osc *OR* 1 = 1	

Table 5-3. A list of Boolean rules with 'oscillation' treated as a 'state'.

5.3.3.3 Determining the attractors of an ESM with more than one sign-inconsistent edge

If two or more sign-inconsistent edges are identified in the ESM, the structural methods described so far encounter several difficulties. After the steady states of the sign-consistent part of the network have been obtained, the order in which the multiple sign-inconsistent edges are put back into the network may possibly affect the obtained attractor. After one sign-inconsistent edge is restored, newly formed sign-inconsistent SCCs might result. As the solving process progresses,

repetitive SCC searching and grouping is required, which is computationally inefficient. Further, the method to determine the attractor based on the positioning of the sign-inconsistent edge relative to the backbone SCC structure is no longer applicable. Within the same SCC, for instance, two sign-inconsistent edges can be positioned such that they are in accordance with each other, both giving rise to a complex attractor (Figure 5-6A), or they can be positioned in discordance forming both motifs in Figure 5-1 (Figure 5-6B). In such cases an algorithm determining the total functionalities of multiple sign-inconsistent edges needs to be constructed in order to obtain the eventual attractor of the network. This method requires intensive computation.

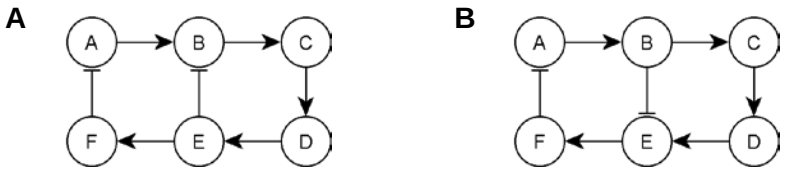


Figure 5-6. Multiple sign-inconsistent edges present in a single SCC. (A) Two negative feedback edges F—|A and E—|B are in accordance with each other. Both cause complex attractors for the SCC, and oscillation will indeed result from (A). Note that the two negative feedback edges do not represent the *minimal* set of sign-inconsistent edges set in this particular case, but they would be in expanded versions that have, e.g., more positive feedforward paths from C to D. (B) Relative to the activation edge C→D, inhibition F—|A forms a negative feedback while inhibition B—|E on the other hand forms an inconsistent feed-forward loop. The final attractor of (B) is an isolation type of steady state.

Therefore, we designed a bottom-up attractor identification algorithm based on state backtracking. Starting from the output node of the network, the algorithm examines the state of the current node (1, 0, or Osc), and generates all viable solutions for the states of its regulator nodes based upon the number of such regulator nodes and the sign of the edges between the regulators and the current node. The states, and henceforth the solution, is tentative until the states

of all nodes in the ESM have been examined and the corresponding solution proved to be self-consistent in terms of nodes' states and consistent with any known external inputs. Otherwise, if any inconsistent nodes states are detected during the process, the tentative solution is discarded. The algorithm resembles the bottom-up algorithm designed for ESM identification but has its own solving mechanism and stopping criterion, which are detailed below.

Assume a node K is currently under examination and has m regulators $\{H_i \mid i = 1,\ldots,m\}$. They are joined together by the *and* operator in the Boolean transfer function of K, as we are inside a single ESM. As a consequence, depending on K's state, a different deduction method applies and different numbers of solutions will be induced:

i) K = 1. Based on the Boolean equalities 1 *and* 1 = 1 and 1 *and not* 0 = 1, all effective inputs of K have to be activations—if the incoming edge to K is positive, the upstream node must be in state 1, and if the edge is negative, the upstream node must be in state 0. A single solution is generated.

ii) K = 0. Based on Boolean equality 0 *and* ANY STATE = 0, the absence of one activator or the presence of one repressor of K is sufficient to put K in the 0 state. m solutions will be generated, with $H_i = 0$ ($H_i = 1$) in the i^{th} solution (i = 1,…,m), if the edge from H_i to K is positive (negative). Meanwhile, H_j can be any state and is henceforth still unsolved for any $j \neq i$.

iii) K is oscillating. As illustrated in Table 5-3, oscillation can also be considered as a special 'state' of a node, and has its own set of Boolean equalities. Therefore, H_i can either be oscillating or activating K (i = 1,…,m), or equivalently H_i = Osc, or $H_i = 1$ ($H_i = 0$) if the edge from H_i to K is positive (negative), which gives us a total number of 2^m solutions, since two choices of states are available for each H_i (i = 1,…,m) regardless of the sign of the edge. The solution in

which none of the H_i is oscillating is not acceptable, for it would result in K = 1 instead of an oscillating K, making the total number of viable solutions under this category ($2^m - 1$).

To initiate the calculation, three separate, tentative solutions are generated with the output node of the ESM in state 1, 0 and Osc, respectively. Subsequently, the bottom up state determination is carried out iteratively until one of the following scenarios occurs: i) The states of all nodes are assigned with no inconsistency encountered. Consequently, the calculation for that particular solution is completed and one attractor of the network is obtained. One can move onto the next tentative solution. ii) A self-consistent partial solution is obtained, indicating that nodes whose states have been solved form a closed loop, with no implication on the states of the rest of the ESM. In such case, a node whose state is not solved yet needs to be initialized with all three possible states again (1, 0, and Osc), generating 3 tentative solutions and the solving process will proceed. iii) One node is required to be in two different states simultaneously via two separate state deductions in the same tentative solution, or in other words, inconsistency is encountered. In this case, the tentative solution in which contradictory states are found is discarded, and one can move on to the next tentative solution.

5.4 References

1. Li, S., Assmann, S. M., and Albert, R. (2006) Predicting Essential Components of Signal Transduction Networks: A Dynamic Model of Guard Cell Abscisic Acid Signaling. *PLoS Biol*. 4, e312.
2. Thakar, J., Pilione, M., Kirimanjeswara, G., Harvill, E. T., and Albert, R. (2007) Modeling Systems-Level Regulation of Host Immune Responses. *PLoSComput Biol*. 3, e109.
3. Ahn, Y., Bagrow, J. P., Lehmann, S. (2010) Link communities reveal multiscale complexity innetworks. *Nature*466, 761–764.
4. Lee, T. I., Rinaldi, N. J., Robert, F., Odom, D. T., Bar-Joseph, Z., Gerber, G. K., Hannett, N. M., Harbison, C. T., Thompson, C. M., Simon, I., Zeitlinger, J., Jennings, E. G., Murray, H. L., Gordon, D. B., Ren, B., Wyrick, J. J., Tagne, J.-B., Volkert, T. L., Fraenkel, E., Gifford, D. K., and Young, R. A. (2002). Transcriptional regulatory networks in *Saccharomyces cerevisiae*. *Science* 298, 799–804.
5. Schwikowski, B., Uetz, P., Fields, S. (2000). A network of protein–protein interactions in yeast.

Nature Biotechnology 18, 1257–1261.
6. Stelzl, U., Worm, U., Lalowski, M., Haenig, C., Brembeck, F. H., Goehler, H., Stroedicke, M., Zenkner, M., Schoenherr, A., Koeppen, S., Tim, J., Mintzlaff, S., Abraham, C., Bock, N., Kietzmann, S., Goedde, A., Toksoz, E., Droege, A., Krobitsch, S., Korn, B., Birchmeier, W., Lehrach, H., Wanker, E. E. (2005) A Human Protein-Protein Interaction Network: A Resource for Annotating the Proteome. *Cell* 122, 957–968.
7. Hatzimanikatis, V., Li, C., Ionita, J. A., Broadbelt, L. J. (2004). Metabolic networks: Enzyme function and metabolite structure. *Curr. Opin. Struct. Biol.* 14, 300–306.
8. Chen, K. C., Csikasz-Nagy, A., Gyorffy, B., Val, J., Novak, B., and Tyson, J. J. (2000) Kinetic analysis of a molecular model of the budding yeast cell cycle. *Mol. Biol. Cell* 11, 369–391.
9. Kauffman, S. A. (1969) Metabolic stability and epigenesis in randomly constructed genetic nets. *J. Theor Biol.* 22, 437–467.
10. Thomas, R. (1973) Boolean formalization of genetic control circuits. *J. Theor Biol.* 42(3), 563–585.
11. Li, F., Long, T., Lu, Y., Ouyang, Q. & Tang, C. (2004) The yeast cell-cycle network is robustly designed. *ProcNatAcadSci USA* 101, 4781–4786.
12. Davidich, M.I., and Bornholdt, S. (2008) Boolean network model predicts cell cycle sequence of fission yeast. *PLoS ONE* 3, e1672
13. Zhang, R., Shah, M. V., Yang, J., Nyland, S. B., Liu, X., Yun, J. K., Albert, R., and Loughran, T. P., Jr. (2008) Network model of survival signaling in large granular lymphocyte leukemia. *ProcNatlAcadSci* USA 105, 16308–16313.
14. 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. *PLoSComput Biol*. 3, e163.
15. Jeong H., Mason S. P., Barabási, A.-L., Oltvai Z. N. (2001) Lethality and centrality inprotein networks. *Nature* 411, 41–42.
16. Albert, R., Jeong, H., Barabási, A.-L. (2000) Error and attack tolerance of complexnetworks. *Nature* 406, 378–382.
17. Stelling, J., Klamt, S., Bettenbrock, K., Schuster, S., Gilles, E. D. (2002) Metabolic networkstructure determines key aspects of functionality and regulation. *Nature* 420, 190–193.
18. Palumbo, M. C., Colosimo, A., Giuliani, A., Farina, L. (2005) Functional essentiality fromtopology features in metabolic networks: a case study in yeast. *FEBS Lett.*579(21), 4642–4646.
19. Shen-Orr, S., Milo, R., Mangan, S., and Alon, U. (2002) Network motifs in the transcriptional regulation network of *Escherichia coli*. *Nature Genetics* 31, 64–68.
20. Thomas, R., D'Ari, R. (1990) "Biological Feedback." 1st ed. CRC Press, Florida.
21. Wang, R., Albert, R. (2011) Elementary signaling modes predict the essentiality of signal transduction network components. *BMC Systems Biol.* 5, 44.
22. DasGupta, B., Enciso, G. A., Sontag, E. D., Zhang, Y. (2007) Algorithmic and complexity results for decompositions of biological networks into monotone subsystems. *Biosystems*90(1), 161–178.
23. Sontag, E. D. (2007) Monotone and Near-Monotone Systems. *Systems and Synthetic Biol.* 1(2), 59–87.
24. Saadatpour, A., Wang, R., Liao, A., Liu, X., Loughran, T. P., Albert, I., Albert, R. (2011) Dynamical and structural analysis of a T cell survivalnetwork identifies novel candidate therapeutic targetsfor large granular lymphocyte leukemia. *PLoS Comput Biol.* 7(11),

e1002267.
25. Veliz-Cuba, A. (2011) Reduction of Boolean network models. *J.Theor Biol*. 289C, 167–172.
26. Naldi, A., Remy, E., Thieffry, D., Chaouiya, C. (2011) Dynamically consistent reduction of logical regulatory graphs. *Theor Comput Sci*, 412, 2207–2218.
27. Smith, H.L. (1995) "Monotone dynamical systems: an introduction to the theory of competitive and cooperative systems". American Mathematical Society, Providence, R.I.
28. Albert, R., DasGupta, B., Hegde, R., Sivanathan, G.S., Gitter, A., Gürsoy, G., Paul, P., Sontag, E. (2011)A new computationally efficient measure of topological redundancy of biological and social networks. *Phys Rev E*, 84(3), 036117.
29. Kaplan, S., Bren, A.,Dekel, E.,Alon, U. (2008)The incoherent feed-forward loop can generate non-monotonic input functions for genes. *Mol. Systems Biology*, 4, 203.
30. Albert, R.,Jeong, H., Barabási, A.-L. (1999)The Diameter of the World-Wide Web. *Nature*, 401, 130-131.
31. Albert, R., Albert, I.,Nakarado, G. L. (2004)Structural Vulnerability of the North American Power Grid.*Phys. Rev. E*, 69, 025103(R).
32. Huang, S., Eichler, G. S., Bar-Yam, Y., Ingber, D. E. (2005) Cell Fates as High-Dimensional Attractor States of a Complex Gene Regulatory Network. *Phys Rev Lett*. 94(12), 128701.
33. Locke, J. C. W., Southern, M. M., Kozma-Bognar, L., Hibberd, V., Brown, P. E.,Turner, M. S., Millar, A. J. (2005) Extension of a genetic network modelby iterative experimentation and mathematical analysis. *MolSystBiol*, 1, 13.
34. Pruneda-Paz, J. L., Kay, S. A. (2010) An expanding universe of circadian networks in higher plants. *Trends Plant Sci*, 15, 259–265.
35. Barabási, A.-L., Oltvai, Z. N. (2004) Network biology: understanding the cell's functional organization. *Nat Rev Genet*, 5(2), 101–113.
36. Rohlf, T., Bornholdt, S. (2002) Criticality in Random Threshold Networks: Annealed Approximation and Beyond. *Physica A*, 310, 245–259.
37. Jorge G.T. Zañudo, Maximino Aldana, Gustavo Martínez-Mekler. (2010) Boolean Threshold Networks: Virtues and Limitations for Biological Modeling. *To appear in a book on Information Processing*.
38. Kim, J.-R., Kim, J., Kwon, Y.-K., Lee, H.-Y., Heslop-Harrison, P., Cho, K.-H. (2011) Reduction of Complex Signaling Networks to a Representative Kernel. *Sci. Signal*, 4(175), 35.
39. Saadatpour, A., Reluga, T., Albert, R. (2012) A network reduction method for Boolean models of biological regulatory networks. *Submitted*.
40. Wang, R., Sun, Z., Albert, R. (2013) Minimal functional routes in directed graphs with dependent edges. *International Transactions in Operational Research* 20, 391-409.

Chapter 6

Conclusions and Outlook

When given a complex system, delineating its components, perusing their regulatory relationship and dynamic interplay and representing the whole with a network is a unique way to study its details without losing the overall picture of the system. In fact, it is often times the only way to generate any information about the system as a whole. With the capacity of modern data acquisition, emergent properties and behaviors on the system level rather than parameterized dynamic details between its elements become the answer to many important questions, giving the approach an increasing popularity among researchers.

In this dissertation, I presented studies focused on network modeling of complex systems with dynamics from both the application of such approach to a signaling process in plants and the theoretical analysis of the dynamics of such models. The studies were conducted with deepening our understanding of the process of stomatal opening, widening the horizon of the applicability of such modeling approach, and exploring the unchartered territory of its dynamic properties in mind. All three goals have been successfully achieved.

By applying the powerful tool of network modeling, extending the methodology of Boolean dynamics to multi-level discrete dynamics, and by organically integrating the two, Chapter 3 presented a successful modeling of the crucial biological process of stomatal opening. The model captured in a qualitative manner a multitude of vital aspects of biological observation, offering sufficient confidence in bringing the model into predicting the behavior of stomata, thus revealing the lack of experimental studies on the regulatory relationship between red light-induced stomatal opening and its inhibition by abscisic acid.

Chapter 4 and 5 delve into analyzing the redundancy and steady state behavior of Boolean networks from a theoretical perspective. By incorporating the functional relations among nodes into the network structure itself, the approach offers a novel measure of network redundancy that can precisely predict the system behavior of single node knockouts. The approach also offers unprecedented opportunity of solving for the steady states of networks with Boolean dynamics.

The study outlined in Chapter 3 has the potential for a variety of expansions and extensions. The model can be expanded to include the photosynthetic and metabolic network cascade on a finer granularity, which can then offer a spectrum of candidate targets of its regulation by ABA. The methodology can be readily applied to systems with similar dynamics. The authors have received multiple requests for its dynamic implementation. Chapter 4 and 5 offer a novel measure of network redundancy/robustness and a standardized, computationally efficient procedure to solve for the attractors of Boolean networks. The redundancy measure precisely predicts for the first time the correlation between the importance of the nodes and the effect of their knockouts without carrying out real time simulations of the network dynamics. Both works have the potential of being extended to beyond Boolean network models.

APPENDIX

Table of system biology terms:

Terms	Definitions
Node	a graphical representation of a component of a biological system, such as a protein, secondary messenger or small molecule.
Edge	any type of interaction that exist between nodes, e.g. chemical reaction, regulation or causal relationship.
Node state	the value of the variable assigned to the node. It can mean concentration, activity, status. In Boolean modeling, there are two node states: ON, meaning above threshold concentration or activity, and OFF, meaning below threshold concentration or activity.
State of the system	an*N*-dimensional vector, where *N* is the total number of nodes in the system. The i*th* dimension is the state of the i*th* node.
Fixed point or steady state	a state vector of the system that is time-invariant. If the system reaches a steady state, it will stay in that state.
State transition graph	a graph showing the evolution of the state of the system. The system will undergo state changes along the direction of the arrows.
Attractor	a set of connected states of the system. Each state in the set has to be reachable by the system after a sufficiently long period of time. Fixed points (steady states) are a subtype of attractors.
Basin of Attraction	a set of connected states of the system. Given sufficiently long time, every state in the set *can* reach a specific attractor of the system. All such states that are able to reach the same attractor form the basin of attraction of *that* particular attractor.

VITA

Zhongyao Sun

EDUCATION

Spring 2015 Ph. D. in Physics, The Pennsylvania State University

(Expected)

2005 B.S. in Physics, Tsinghua University

RESEARCH INTEREST

Graph/network theory, network modeling, network dynamics, computational methods based on networks.

HONORS AND AWARDS

2009, 2007, 2005 Duncan Fellowship, The Pennsylvania State University, Department of Physics

PROFESSIONAL EXPERIENCE

Graduate Research Assistant in Physics, The Pennsylvania State University, University Park, PA

(2006 - Present)

Teaching assistant for Physics 212 (Introductory Electrodynamics) and Physics 251 (Introductory Electrodynamics), The Pennsylvania State University, University Park, PA

(2007-2008, 2015)

PUBLICATIONS AND MANUSCRIPTS

Sun, Z., Jin, X., Albert, R., Assmann, S.M. (2014) Multi-level modeling of light-induced stomatal opening offers new insight into its regulation by drought. *PLOS Computational Biology* 10, e1003930

Wang, R., Sun, Z., Albert, R. (2013) Minimal functional routes in directed graphs with dependent edges. *International Transactions in Operational Research* 20, 391-409

Sun, Z., Albert, R. node-independent elementary signaling modes: a measure of redundancy in Boolean signaling transduction networks. *Manuscript in preparation.*