

# **Collective dynamics and pattern formation in systems of communicating cells**

Theory and Simulations

## **Proefschrift**

ter verkrijging van de graad van doctor  
aan de Technische Universiteit Delft,  
op gezag van de Rector Magnificus prof. dr. ir. T.H.J.J. van der Hagen,  
voorzitter van het College voor Promoties,  
in het openbaar te verdedigen op woensdag 8 april 2020 om 10:00 uur

door

**Yiteng DANG**

Master of Advanced Study in Mathematics, University of Cambridge,  
Verenigd Koninkrijk,  
Master of Science in Quantum Physics, École Normale Supérieure, Frankrijk  
geboren te Wuxi, China.

Dit proefschrift is goedgekeurd door de

promotor: prof. dr. A.M. Dogterom

copromotor: dr. H.O. Youk

Samenstelling promotiecommissie:

Rector Magnificus,  
Prof. dr. A.M. Dogterom  
Dr. H.O. Youk

voorzitter  
Technische Universiteit Delft  
Technische Universiteit Delft

*Onafhankelijke leden:*

Prof. dr. Y.M. Blanter	Technische Universiteit Delft
Prof. dr. J. Garcia Ojalvo	Universitat Pompeu Fabra
Prof. dr. ir. T.E.A. de Greef	Technische Universiteit Eindhoven
Prof. dr. R.M.H. Merks	Universiteit Leiden
Prof. dr. N.H. Dekker	Technische Universiteit Delft, reservelid



*Keywords:* pattern formation, multicellular systems, cell-cell communication, self-organization, complex systems, cellular automata, gene networks

*Printed by:* Ipskamp Printing, Enschede

*Front & Back:* Yiteng Dang

Copyright © 2020 by Y. Dang

Casimir PhD Series, Delft-Leiden 2020-08

ISBN 978-90-8593-435-6

An electronic version of this dissertation is available at  
<http://repository.tudelft.nl/>.

# Contents

<b>Summary</b>	<b>vii</b>
<b>Samenvatting</b>	<b>x</b>
<b>I Introduction</b>	<b>1</b>
<b>1 Preface</b>	<b>2</b>
1.1 Self-organization in biology . . . . .	2
1.2 Biological pattern formation: a concise history . . . . .	3
1.3 Scope and research questions of this thesis . . . . .	5
1.4 Structure of this thesis . . . . .	7
<b>2 Extended introduction: Pattern formation in multicellular systems</b>	<b>9</b>
2.1 Cell-cell communication: an overview . . . . .	9
2.2 Self-organization in multicellular systems . . . . .	15
2.3 Models and mechanisms of multicellular pattern formation . . . . .	21
<b>II Collective dynamics of communicating cells: modeling and analysis</b>	<b>30</b>
<b>3 Cells communicating with one signaling molecule generate diverse collective behaviors</b>	<b>31</b>
3.1 Multi-scale model for quorum sensing cells . . . . .	31
3.2 Synchronization and autonomy in communicating analogue cells . . . . .	35
3.2.1 Modeling cells with continuous response curves . . . . .	36
3.2.2 Collectiveness versus autonomy . . . . .	38
3.2.3 Exact solutions of a one-dimensional model . . . . .	43
3.2.4 Transition from collective to autonomous dynamics . . . . .	44
3.2.5 Persistence of bistability . . . . .	51

3.2.6	Discussion . . . . .	55
3.2.7	Appendix . . . . .	58
3.3	Sensitivity to initial conditions . . . . .	62
3.3.1	Discrete system (infinite Hill coefficient) . . . . .	62
3.3.2	Continuous system (finite Hill coefficient) . . . . .	65
3.3.3	Discussion . . . . .	69
3.4	Using mutual information to characterize pattern formation . . . . .	70
3.4.1	Predicting and inferring initial and final states . . . . .	70
3.4.2	Continuous system . . . . .	75
3.4.3	Application: ensemble-level responses to stochastic noise . . . . .	78
3.4.4	Discussion . . . . .	81
<b>4</b>	<b>Statistical dynamics of spatial-order formation by communicating cells</b>	<b>83</b>
4.1	Introduction . . . . .	84
4.2	Results . . . . .	85
4.2.1	Cellular automaton simulates secrete-and-sense cells that slowly respond to a rapidly diffusing signaling molecule. . . . .	85
4.2.2	Secrete-and-sense cells can be classified into distinct behavioral phases. . . . .	86
4.2.3	Grouping multiple spatial configurations into one macrostate based on their common spatial index $I$ and fraction $p$ of cells that are ON. . . . .	87
4.2.4	Cellular lattice is represented by a particle whose position $(p, I)$ and trajectory depends on the behavioral phase. . . . .	88
4.2.5	Cellular lattice acts as a particle that rolls down on and adheres to a pseudo-energy landscape . . . . .	92
4.2.6	Gradient of the pseudo-energy and the trapping probability $P_{eq}(p, I)$ completely specify the particle's motion . . . . .	93
4.2.7	Stochastic sensing can yield spatial configurations that are more ordered than those formed without noise. . . . .	95
4.3	Discussion . . . . .	98
S4.4	Supplemental Data Items . . . . .	102
S4.5	Supplementary Information . . . . .	106
S4.5.1	Properties of the spatial index $I$ . . . . .	106
S4.5.2	Deriving the maximum allowed value of $ I $ for each $p$ . . . . .	109
S4.5.3	Density of states as a function of $(p, I)$ . . . . .	110
S4.5.4	Proof that the pseudo-energy $h$ is a non-increasing function over time . . . . .	113
S4.5.5	Derivation of the trapping probability — $P_{eq}(p, I)$ . . . . .	114

S4.5.6	Equation of motion derived from the pseudo-energy . . . . .	116
<b>5</b>	<b>Cellular dialogues that enable self-organization of dynamic spatial patterns</b>	<b>124</b>
5.1	Introduction . . . . .	125
5.2	Results . . . . .	127
5.2.1	Computational search for cellular dialogues that enable self-organized patterns . . . . .	127
5.2.2	Cellular dialogues enable self-organization of wide array of dynamic patterns . . . . .	128
5.2.3	Common structural elements in cellular dialogues that generate dynamic patterns . . . . .	130
5.2.4	Analytic framework explains how cells collectively sustain dynamic spatial patterns . . . . .	134
5.2.5	Self-organization occurs through a three-stage, “order-fluctuate-settle” mechanism . . . . .	137
5.2.6	Dynamic patterns with more complex elements . . . . .	143
5.3	Discussion . . . . .	145
S5.4	Methods . . . . .	148
S5.4.1	Detailed description of our model . . . . .	148
S5.4.2	Simulation and analysis of the model . . . . .	158
S5.4.3	Extending the model by adding complex elements . . . . .	164
S5.5	Supplementary Information . . . . .	170
S5.5.1	Overview of self-organized patterns . . . . .	170
S5.5.2	Parameter-derived general constraints on the dynamics . . . . .	176
S5.5.3	Analytic framework for traveling wave propagation . . . . .	188
S5.5.4	Robustness and reliability of traveling waves . . . . .	210

### **III Epilogue** **227**

<b>6</b>	<b>Discussion</b>	<b>228</b>
6.1	Key lessons learned . . . . .	228
6.2	Relation to other mathematical models . . . . .	232
6.3	Biological realism of our models . . . . .	234
6.4	Experimental validation . . . . .	236
6.4.1	Developmental biology systems . . . . .	237
6.4.2	Synthetic biological systems . . . . .	238
6.4.3	Non-biological synthetic systems . . . . .	240
6.5	Directions for future research . . . . .	241

<b>Acknowledgements</b>	<b>250</b>
<b>List of Figures</b>	<b>253</b>
<b>References</b>	<b>256</b>
<b>Curriculum Vitæ</b>	<b>274</b>
<b>List of Publications</b>	<b>275</b>

# SUMMARY

How a system of genetically identical biological cells organizes into spatially heterogeneous tissues is a central question in biology. Even when the molecular and genetic underpinnings of cell-cell interactions are known, how these lead to multicellular patterns is often poorly understood. Of particular interest are dynamic patterns such as traveling waves, which confer spatiotemporal control over key developmental processes such as differentiation, segmentation and cell division. Theoretical approaches based on mathematical descriptions of underlying physical and chemical processes provide a promising avenue to explore biological pattern formation. In particular, theoretical models connect processes on the molecular scale to biological function on the tissue level and may provide mechanistic descriptions of how patterns are generated and maintained.

Pattern formation requires coordination of chemical, physical and biological processes at different length and time scales. Classical models based on reaction-diffusion mechanisms or morphogen gradients typically focus on chemical processes with a single scale parameter, and have no explicit treatment of cells and gene regulatory networks. To overcome these limitations, we developed multiscale, agent-based models to study the gene expression dynamics of cells that secrete and sense signaling molecules. Our models explicitly incorporate cellular processes by modeling cells as interacting agents whose response is dictated by realistic descriptions of underlying biochemical and gene regulatory processes.

With our theoretical model, we observed a variety of population-level spatial and temporal patterns that formed without external morphogen gradients. A central question arising from this observation is whether there are “rules” at the microscopic level (e.g. related to gene regulatory networks and molecular parameters) that govern features of the macroscopic behavior (i.e. gene expression patterns). We addressed this question by performing large-scale simulations of our models and developing theoretical frameworks to characterize and explain these patterns.

We first studied cells that communicate with a single signaling molecule, using concepts and approaches from dynamical systems theory, cellular automaton theory and information theory. We extended previous models by considering “analogue” cells that respond to sensed concentrations in a continuous way, rather than “digital” cells that have

a binary response (i.e. either turn “ON” or “OFF” a gene). We observed that this system shows a transition between an “autonomous” and a “synchronized” state and proposed various quantitative measures to characterize this transition. We also probed the system’s sensitivity to initial conditions by defining and quantifying perturbations and responses. Finally, we proposed a metric for characterizing the cell population’s degree of autonomy based on the information-theoretic concept of mutual information.

Next, we constructed a macroscopic description for the same model with one signaling molecule, which describes how population-level features of the system evolve. We did this by focusing on quantities that characterize population level features without specifying the state of each cell. These quantities characterize the average gene expression level and spatial correlations between different cells. We then derived a “pseudo-energy” based on an analogy with spin models in statistical physics. This is a phenomenological quantity that monotonically decreases over time, and can be expressed solely in terms of these macroscopic quantities. This revealed a conceptual picture of our multicellular system as a particle rolling down a “pseudo-energy” landscape. Based on these facts, we derived an “equation of motion” that independently describes the evolution of these macroscopic quantities and largely follows the gradient of the landscape. This “equation of motion” recapitulates main qualitative findings and gives a visual interpretation for why the system becomes increasingly spatially ordered over time.

We subsequently extended our theoretical model to cells that communicate with multiple signaling molecules. With two types of molecules, we observed self-organized dynamic patterns such as collective oscillations and traveling waves. As there are multiple ways in which these two molecules can regulate each other at the genetic level, a primary question is how these different ways of communicating relate to the various emergent multicellular patterns. To address this question, we first performed a computational search that enumerated all two-gene networks and scanned a large range of parameters, to find the various conditions under which the different types of patterns could form. This revealed that the various ways of communicating with two molecules grouped into three classes, which are characterized by easy-to-state rules regarding their underlying network topologies. One of these classes is capable of generating dynamic spatial patterns, such as traveling waves and spiral waves, which can propagate information across the system. We then constructed a theoretical framework to study traveling waves. In essence, we derived mathematical conditions for the propagation of these waves in terms of the system’s microscopic parameters. This theoretical approach recovers the same ways of communicating (i.e. gene networks for two signaling molecules) that enable traveling waves to form as found in simulations, and correctly retrieves the corresponding parameter conditions. Furthermore, we found that complex dynamic

patterns in our model emerge through a three-stage process. Starting from random initial conditions, the system quickly becomes spatially ordered, after which transient incoherent waves traverse the system in a turbulent fashion, until they suddenly settle into dynamic patterns that repeat themselves over time without changing shape. Finally, we extended our model to include other realistic features of biological tissues (e.g. stochasticity) and quantitatively characterized their impact on the formation and propagation of dynamic spatial patterns.

Altogether, our results show that features at the molecular and subcellular scales directly affect pattern formation in multicellular systems. This suggests that the collective dynamics of multicellular systems and tissue-level patterns can be controlled by appropriately selecting gene regulatory architectures and tuning molecular parameters. Furthermore, our work has wider implications for understanding and engineering biological systems from experimental and theoretical perspectives. Experimentally, our results provide a blueprint for identifying mechanisms and conditions enabling pattern formation in multicellular systems (e.g. in development), as well as for engineering patterns in synthetic biological systems or even non-living chemical systems (e.g. artificial protocells). Theoretically, the methods we have developed can be extended to study more specific biological systems. Furthermore, they may offer insights into complex dynamics in other types of discrete interacting systems, in biology and beyond.

# SAMENVATTING

Hoe een systeem van genetisch identieke biologische cellen zich ontwikkelt tot ruimtelijk heterogene weefsels is een hoofdvraag in de biologie. Kennis van de onderliggende moleculaire en genetische interacties binnen en tussen cellen leidt niet direct tot begrip van hoe multicellulaire patronen ontstaan. In het bijzonder zijn dynamische patronen zoals lopende golven belangrijk, omdat deze ruimtelijke controle over belangrijke ontwikkelingsprocessen verschaffen, zoals differentiatie, segmentatie en celdeling. Theoretische benaderingen gebaseerd op wiskundige beschrijvingen van onderliggende fysische en chemische processen vormen een veelbelovende methode om biologische patroonvorming te verkennen. In het bijzonder verbinden theoretische modellen processen op moleculaire schaal met biologische functie op weefselniveau en kunnen ze mechanistische beschrijvingen geven van hoe patronen vormen en in stand worden gehouden.

Patroonvorming vereist coördinatie van chemische, fysische en biologische processen op verschillende lengte- en tijdschalen. Klassieke modellen gebaseerd op reactie-diffusie mechanismen of morfogengradiënten richten zich meestal op chemische processen met een enkele schaalparameter en hebben geen expliciete beschrijving van cellen en genetische netwerken. Om deze tekortkomingen tegemoet te komen hebben we multischaal, agent-gebaseerde modellen ontwikkeld om de genexpressiedynamica te bestuderen van cellen die signaalmoleculen vrijlaten en waarnemen. Door cellen te modelleren als interacterende individuen, kunnen we cellulaire processen expliciet beschrijven. Verder modelleren we de respons van de cellen door middel van realistische beschrijvingen van de onderliggende genregulatie en andere biochemische processen.

Met ons theoretisch model hebben we verschillende patronen op weefselniveau waargenomen, die zonder externe morfogenen tot stand zijn gekomen. Deze omvatten zowel ruimtelijke patronen als patronen die in de tijd regulier gedrag vertonen. Deze observatie leidt tot de vraag of er op microscopisch niveau “regels” zijn (bijv. eigenschappen van genetische netwerken en moleculaire parameters) die kenmerken van het macroscopische gedrag reguleren (d.w.z. genexpressiepatronen). Om deze vraag te beantwoorden hebben we grootschalige simulaties van onze modellen uitgevoerd en theoretische kaders ontwikkeld om de patronen te karakteriseren en te verklaren.

We hebben eerst cellen bestudeerd die communiceren met een enkel signaalmolecuul. Hiervoor hebben we concepten en methodes gebruikt uit dynamische systeemtheorie, cellulaire automatentheorie en uit de informatietheorie. We hebben eerder ontwikkelde modellen uitgebreid door analoge cellen in plaats van digitale cellen te nemen. Dat wil zeggen dat we cellen modelleren die als respons op waargenomen concentraties hun genexpressie niet op binaire wijze reguleren (door een gen “aan” of “uit” te zetten), maar op continue wijze. Dit systeem vertoont een overgang tussen een “autonome” en een “gesynchroniseerde” fase, die we kunnen karakteriseren door middel van verschillende grootheden die we hebben geïntroduceerd. We hebben ook de invloed van beginwaarden onderzocht door perturbaties en responsen te definiëren en te kwantificeren. Tot slot hebben we een grootheid geïntroduceerd om de mate van autonomie van de celpopulatie te meten die gebaseerd is op een concept uit de informatietheorie genaamd transinformatie.

Vervolgens hebben we een macroscopische beschrijving voor hetzelfde model met één signaalmolecuul geconstrueerd, die beschrijft hoe eigenschappen op populatieniveau zich ontwikkelen in de tijd. Hiervoor hebben we grootheden gebruikt die het systeem op populatieniveau karakteriseren zonder de toestand van elke cel te specificeren. Voorbeelden zijn het gemiddelde genexpressieniveau van de populatie en de ruimtelijke correlaties tussen verschillende cellen. We hebben daarna een “pseudo-energie” afgeleid op basis van een analogie met spinsystemen uit de statistische fysica. Dit is een fenomenologische grootheid die monotoon afneemt en direct uit te drukken is in deze macroscopische grootheden. Hieruit ontstond een conceptueel beeld van ons meercellige systeem als een deeltje dat op een landschap van deze “pseudo-energie” naar beneden rolt. Op basis van deze bevindingen hebben we een “bewegingsvergelijking” afgeleid voor deze macroscopische grootheden, die grotendeels de gradiënt van het landschap volgt. Deze “bewegingsvergelijking” vat de belangrijkste kwalitatieve bevindingen van het model samen en geeft een visuele interpretatie van het toenemen van ruimtelijke orde in het systeem.

Daarna hebben we ons theoretisch model uitgebreid door cellen te bestuderen die communiceren met meerdere signaalmoleculen. Met twee soorten moleculen hebben we zelfgeorganiseerde dynamische patronen geobserveerd, zoals collectieve oscillaties en lopende golven. Aangezien er meerdere manieren zijn waarop deze twee moleculen elkaar op genetische niveau kunnen reguleren, is een primaire vraag hoe deze verschillende manieren van communiceren zich verhouden tot de verschillende emergente meercellige patronen. Om deze vraag te beantwoorden, hebben we eerst een computeralgoritme geïmplementeerd, die voor elk van de netwerken met twee signaalmoleculen een groot aantal parameters doorzoekt, om condities te vinden waaronder de verschillende

soorten patronen kunnen vormen. Hieruit bleek dat de verschillende wijzen waarop cellen met twee signaalmoleculen kunnen communiceren groeperen in drie klassen, die worden gekenmerkt door simpele regels met betrekking tot hun netwerktopologieën. Eén van deze klassen is in staat om dynamische ruimtelijke patronen te genereren, zoals lopende golven en spiraalvormige golven, waarmee cellen op grote schaal informatie met elkaar kunnen uitwisselen. Daarom hebben we een theoretisch kader geconstrueerd om lopende golven te bestuderen. In essentie hebben we wiskundige vergelijking afgeleid voor de propagatie van deze golven, die we kunnen uitdrukken in termen van de microscopische parameters van het systeem. Met deze theoretische benadering hebben we dezelfde communicatievormen gevonden die in staat zijn om lopende golven te vormen (d.w.z. gennetwerken met twee moleculen) als in onze simulaties, met vrijwel identieke condities op de parameters. Verder hebben we geconstateerd dat complexe dynamische patronen in ons model ontstaan in drie fasen. Met willekeurige begintoestanden wordt het systeem snel ruimtelijk geordend, waarna incoherente lopende golven het systeem op turbulente wijze doorkruisen, totdat deze plotseling coherent worden en dynamische patronen vormen die zich herhalen in de tijd zonder van gedaante te veranderen. Tenslotte hebben we ons model uitgebreid door andere realistische kenmerken van biologische weefsels (bijv. stochasticiteit) mee te nemen en de invloed hiervan op de vorming en instandhouding van dynamische ruimtelijke patronen kwantitatief te karakteriseren.

Al met al laten onze resultaten zien dat moleculaire en subcellulaire eigenschappen van cellen directe invloed hebben op patroonvorming in meercellige weefsels. Dit suggereert dat de collectieve dynamica van multicellulaire systemen en patronen op weefselschaal kan worden afgestemd door specifieke netwerkarchitecturen en moleculaire parameters te kiezen. Verder heeft ons werk bredere implicaties voor het begrijpen en nabouwen van biologische systemen vanuit zowel experimenteel als theoretisch perspectief. Experimenteel bieden onze resultaten een blauwdruk voor het identificeren van mechanismen en omstandigheden die patroonvorming mogelijk maken in meercellige systemen (bijv. in de ontwikkelingsbiologie), evenals voor het bouwen van synthetisch biologische of zelfs niet-levende chemische systemen (bijv. kunstmatige protocellen). Theoretisch kunnen de methodes die we hebben ontwikkeld worden uitgebreid en toegepast om meer specifieke biologische systemen te bestuderen. Bovendien kunnen ze inzichten bieden in complexe dynamica van andere discrete interacterende systemen, zowel in de biologie als daarbuiten.

# I

## INTRODUCTION

# 1

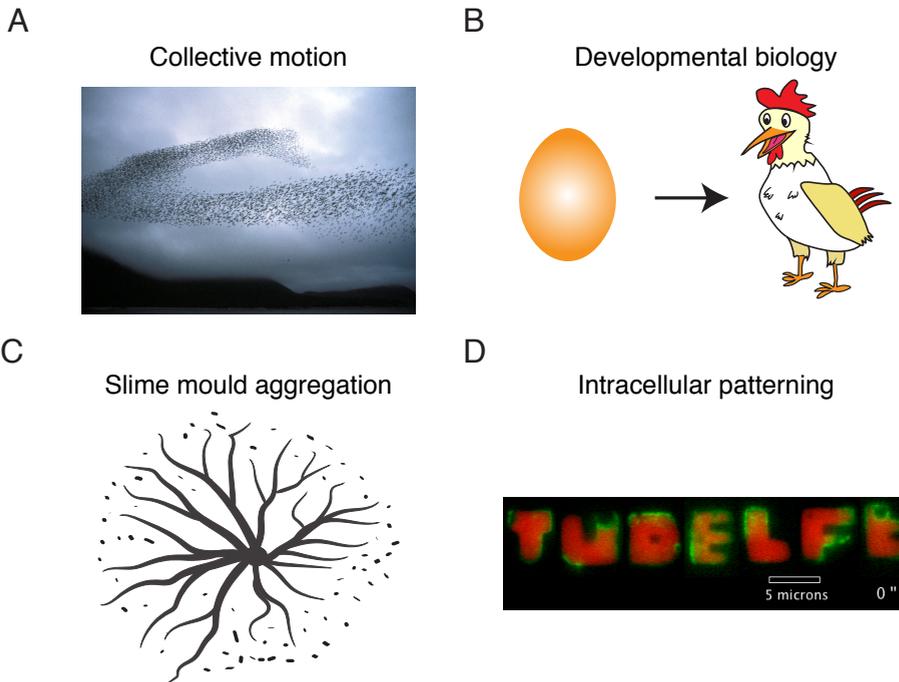
## PREFACE

### 1.1. SELF-ORGANIZATION IN BIOLOGY

Collective phenomena and self-organization are ubiquitous in biology. At the largest scale, the flocking of birds and the schooling of fish show how communicating organisms coordinate their behavior (Fig. 1A). At the scale of tissues, cells interact chemically and mechanically to form patterns, often as a first step in a complex program of development that morphs a single fertilized egg cell into a multicellular organism (Fig. 1B). In the world of microbes, the slime mould *Dictyostelium discoideum* happily lives as isolated single cells when food is abundant, but the cells aggregate into a colony when food is scarce (Fig. 1C). Finally, at the intracellular scale, proteins form dynamic patterns such as oscillations and waves to coordinate processes such as cell division (Fig. 1D).

The focus on this thesis is on self-organization at the scale of multicellular systems such as tissues. Before delving into the details, let us explicitly discuss two conceptual ideas, because they form the backbone that holds this thesis together. First, there is the concept of *self-organization*. Cells are able to accomplish tasks collectively that they would not accomplish if they all acted individually. Yet in many multicellular systems, the cells are genetically identical and there is no natural “leader cell” orchestrating the whole process. The collective behavior of the cells thus results from the interactions between individual cells, rather than from manipulation by an external agent. These interactions result in *emergent phenomena* such as collective immune response or aggregation of cells. They cannot be explained by examining the individual cells making up the collective. In Aris-

totle's words, "the whole is more than the sum of its parts".



**Figure 1.1:** Biological examples of self-organization. (A) Bird flocking arises without “leader birds” orchestrating the process — rather, each bird aligns its motion in response to how its neighbors are flying to avoid collision. Image courtesy of Jan van IJken. (B) The development of an organism from embryo to full adult requires many processes that are not directed by external factors, but self-organize as a result of underlying genetic processes. (C) Aggregation of slime mold *Dictyostelium discoideum* is a result of cAMP signaling between cells. (D) Protein patterns in *E. coli* form patterns in response to cell geometry. Image courtesy of Cees Dekker lab.

## 1.2. BIOLOGICAL PATTERN FORMATION: A CONCISE HISTORY

The work presented in this thesis follows in the footsteps of numerous great minds and influential ideas, both directly through the concepts and methods they have developed, as well as indirectly through their research philosophies. Here, we mention three historically significant scientific developments that have shaped the present work.

As arguably the founding father of modern theoretical biology, D’Arcy Thompson was one of the pioneers in proposing the idea that physical laws can be used to understand biological form, thereby dispelling previous ideas about *vitalism*, which proposes that living organisms are fundamentally different from inanimate material. Moreover, Thompson proposed that simple mathematical ideas could explain seemingly complicated biological phenomena. In his seminal book *On Growth and Form*, he proposes mathemati-

cal laws to explain a variety of natural patterns and forms, including spirals, leaf arrangements (phyllotaxis), cells and tissues [Thompson, 1917]. However, despite the impressive size and popularity of the work, the proposed mathematical laws do not by themselves constitute causal explanations for biological form [Briscoe and Kicheva, 2017]. Moreover, the role of genetics and evolution are largely neglected, and even rejected by Thompson.

More than three decades after Thompson, the mathematician and polymath Alan Turing — who was well aware of Thompson’s work — proposed a novel theory for patterns in development based on diffusing chemicals. Whereas Thompson’s theories use geometry and mechanics to explain various patterns in biology, Turing’s work invokes self-organizing chemical patterns and is more focused on explaining patterns in developmental biology. Specifically, he demonstrates using elegant mathematical derivations how simple systems of a few chemical species that diffuse and react with each other — now known as *reaction-diffusion systems* — can self-organize into regularly spaced patterns such as stripes and spots [Turing, 1952]. He also coined the term “morphogen” for describing the diffusing chemicals. We will review the recent progress and limitations of Turing’s framework in more detail in Chapter 2.

A third major influence on this work comes from the study of cellular automata. These are spatial models with discrete time, space and states that describe how a collection of “cells” (pixels) change their state over time in response to their local environment (i.e. the states of the neighbors). The dominant driving force behind using cellular automata as models for complexity is Stephen Wolfram [Wolfram, 2002]. While cellular automata are simple models that can be easily simulated on any modern computer, the range of complex patterns that they can produce with very simple rules is astounding. As such, they can be regarded as prototype models for studying how complexity arises from systems with simple underlying rules. One particular class of cellular automata studied by Wolfram are so-called elementary cellular automata, where cells interact only with their immediate neighbors. Wolfram categorized all elementary cellular automata in terms of their dynamic behavior, and showed that a particular limited subset of them form highly complex, chaotic patterns. In essence, the model we introduce later is also a cellular automaton, but we couple it to reaction-diffusion equations. Furthermore, in Chapter 5, we also aim to classify all possible models within our framework, where the degree of freedom we vary is not the local interaction rule for the cellular automaton, but the underlying genetic network governing the interactions between signaling molecules.

Altogether, the work presented in this thesis follows in the tradition of employing mathematical and physical concepts to understanding biological form. Although we rarely directly invoke or utilize any results from these early works, their lasting influence manifests through the approach we take as well as in the research questions we propose, as

will be illustrated in the next section.

### 1.3. SCOPE AND RESEARCH QUESTIONS OF THIS THESIS

The focus of this thesis is on self-organization in populations of cells that communicate chemically through producing and responding to signaling molecules. As such, we have taken an approach which mirrors Turing's approach – by trying to derive general results from simple, well-established principles of chemical signaling, but at the same time leaving out other interactions (e.g. mechanical forces) for the sake of not making the system too complicated. However, more than half a century after Turing, we have a far better understanding of biological pattern formation and need to incorporate this knowledge into the models we build today. For example, since many processes take place at different time and length scales, this calls for multiscale models that combine elements at different time and length scales.

The main questions we addressed are essentially the same questions that Turing and many others raised decades ago, because they are still relevant today. How do patterns and structure emerge from interactions at the molecular level (e.g. between genes inside a cell)? How does an initially isotropic biological system break its symmetry? What are mechanisms of generating heterogeneity in a population of identical cells?

We addressed these questions by constructing and analyzing mathematical models of communicating cells. Our starting point is a cell-based multiscale model developed earlier in our group [Maire & Youk, 2015a]. In essence, this model describes the collective dynamics of a group of chemically communicating cells in terms of the time evolution of their gene expression states. From a theorist's perspective, this raises a number of questions which are common to other studies of complex systems. Can we derive a higher-level description of the system, so that we can forget about individual cells, but still make accurate predictions? How do we classify the emergent phenomena and summarize the conditions under which each of them can arise? Are there conserved quantities, or principles mirroring free energy minimization in our system?

Given the enormous body of literature on pattern formation, self-organization and complex systems, how can we possibly come up with a unique and novel approach to studying a century-old problem? Let me mention two features which distinguish our approach and set it apart from the majority of studies in the existing literature. First, we explore cellular automaton modeling which explicitly incorporates relevant biophysical phenomena – such as secretion and diffusion of signaling molecules – rather than using phenomenological rules for describing how the system should evolve. Cellular automata and other discrete models have been used to model biological pattern formation in a wide range of systems [Deutsch & Dormann, 2005]. However, the existing studies con-

concentrate mostly on lattice gas automata to simulate diffusion and fluid flow, for which rigorous mathematical approaches exist. However, analytical approaches for these models, such as the lattice Boltzmann methods (discussed in [Deutsch & Dormann, 2005]), cannot be directly extended to models of biological cells. This calls for the development of new mathematical tools to analyse cellular automaton models of biological pattern formation.

Secondly, in modeling complex systems one has to make a choice between realism and tractability of the model. We may be tempted to include as many elements as possible into a model, to get an accurate representation of the system. However, by doing so, we often lose the ability to fully comprehend our model because of various reasons: the number of parameters becomes too large, we can no longer deduce steady states and other properties without running brute-force simulations, and even if we could “solve” the entire model, the high dimensionality of the system would still make it hard for us humans to fully comprehend it. Thus, there is a trade-off between “model realism” – how closely the model resembles the actual system – and “model tractability” – how well we can understand the model by analyzing its outcomes. In this work, we have aimed to strike a balance between the two, by studying a model that is complicated enough to incorporate realistic biological features, while at the same time simple enough to allow for the employment of computational tools for analysis.

Our research strategy consists of two parallel paths, which we will traverse simultaneously throughout this thesis. On the first track, we extended a previously developed model [Maire & Youk, 2015a] to include more realistic elements of communicating multicellular systems. The goal here is not merely to add more detail, but to explore what new phenomena arise at the population-level (i.e. in terms the patterns that form) as one adds more complex elements. We also examined whether and how findings of the original model altered as we included these more complex elements. This allowed us to assess the relative importance of various features for generating particular types of behavior. For instance, suppose we observe that a certain gene network is capable of generating traveling waves. We can then ask which elements of this network are most essential to performing this function (of generating waves). If we add or remove interactions between genes, will the network still generate waves?

The other path is more theory-focused and deals with the development of theoretical methods to analyze, interpret and predict simulation results. Once we have defined a model, we could run many simulations of the model and examine by eye what happens in the system in each scenario. However, this would only give us qualitative information such as whether a pattern forms or not, and becomes unfeasible as the system becomes more complex. Therefore, we ran large sets of simulations with different parameters

and initial conditions, and devised automated procedures for analyzing the simulation results and to obtain relevant statistical quantities from these batch simulations. Identifying relevant statistical features was a major challenge in this approach. Once we developed an automated analysis workflow, several questions become simple to address. One key question was how to connect microscopic features of the model (i.e. at the level of the cells) with macroscopic observations (i.e. at the population level). For instance, we may observe that a particular spatial pattern only occurs for a particular set of gene circuits for the signaling molecules. We then sought to explain this observation by deriving a mathematical framework that links the microscopic rules of the system to the large-scale dynamics.

Addressing our research questions required the invention of new tools and concepts specifically tailored to multicellular systems. It may be tempting to take the reductionist's approach and arrogantly reduce all of biology to physics, but it is unlikely that this will give us a better understanding of biological phenomena. As theoretical physicist Philip Anderson remarked, "*At each level of complexity entirely new properties appear, and the understanding of the new behaviors requires research which I think is as fundamental in its nature as any other. [...] Psychology is not applied biology, nor is biology applied chemistry*" [Anderson, 1972].

## 1.4. STRUCTURE OF THIS THESIS

**Chapter 2** is an extended introduction to self-organization and pattern formation in multicellular systems. First, we discuss different forms of cell-cell communication that enable cells to confer information to one another. This enables the emergence of collective phenomena in multicellular systems, and we zoom in on a few examples. Traditionally, pattern formation is explained through a handful of mechanisms and mathematical models. We discuss the validity and limitations of these models.

**Chapters 3-5** cover the bulk of this thesis and present in-depth theoretical analysis of different variants of a mathematical model for communicating cells. **Chapter 3** starts by introducing this model in its simplest form. We then study a particular realization of this model, for which we find that the system transitions between a "synchronous phase" and an "autonomous phase". We then study the sensitivity to initial conditions of this model by performing local perturbations and characterizing how they propagate. Finally, we propose a measure for characterizing the degree of autonomy, based on the mutual information between input and output states of the system.

In **Chapter 4**, we introduce a framework for studying the macroscopic dynamics of our system, by exploiting an analogy of our model with well-known models in statistical physics. Specifically, we define macrostates for our system and derive a "pseudo-energy"

landscape that predicts how the macrostates evolve over time. Together, this yields an equation of motion for the macrostates that recapitulates main features of the macroscopic dynamics.

In **Chapter 5**, we extend our framework to examine cells that communicate through multiple signaling molecules and show that various novel types of patterns emerge, such as complex oscillations and traveling waves. We then find the gene circuits, specifying the interaction between two types of signaling molecules, that enable these patterns to form. We characterize the self-organization of these patterns as a multi-step process. We then introduce more complex elements that correspond to various biological features not taken into account in our original model.

Finally, in **Chapter 6**, we summarize our main findings, discuss limitations of our approach and provide suggestions for future research.

# 2

## EXTENDED INTRODUCTION: PATTERN FORMATION IN MULTICELLULAR SYSTEMS

### 2.1. CELL-CELL COMMUNICATION: AN OVERVIEW

Just like humans have different ways of communicating, cells also have a variety of ways through which they send and receive signals. Cell signaling often refers to *chemical signaling* — the production, secretion and uptake of small molecules that can encode information relevant to other cells in the system. However, cells also interact through *mechanical signaling*, mediated by direct interaction between cells as well as indirectly through interaction of the cells with the extracellular matrix. In recent years several other forms of signaling have been discovered, including transport through tunneling nanotubes and electrical signaling in biofilms. In the following, I will discuss these types of cell signaling by focusing on the mechanisms through which they work, and illustrate their effects through examples. While the rest of the thesis primarily addresses chemical signaling, the other types of signaling are mentioned for completeness and may provide a source of inspiration for future modeling attempts.

## CHEMICAL SIGNALING

The most frequently encountered and arguably best studied form of cellular communication is chemical signaling [Alberts et al., 2007]. In essence, chemical signaling revolves around the production, secretion and uptake of signaling molecules and the resulting cellular response. While there is considerable variety in the biochemical pathways and intracellular machinery that enable chemical signaling, the net effect can often be summarized as the upregulation or downregulation of one or more target genes. If this gene encodes for a signaling factor, then there is feedback between what a cell senses and what it secretes.

Furthermore, chemical signaling takes place on a range of different time and length scales. Stimuli can reach the brain in a fraction of a second and induce immediate reflexes, whereas differentiation of stem cells into somatic cells — a process that is often highly dependent on environmental cues — may take days or even weeks. Signaling can be local, affecting only the directly vicinity of a cell, or it can affect cells far away, for instance in a different part of the body of a multicellular organism.

These differences are described by different modes of cellular communication [Doğaner et al., 2016]. At the closest scale, *juxtacrine* or *contact-dependent signaling* occurs only when two cells physically touch each other (Figure 2.1A). A famous example of a contact-dependent signaling system is the Notch signaling pathway, where two membrane proteins Notch and Delta bind to each other to transmit usually inhibitory signals. Delta binds to the Notch receptor and induces a series of cleavage steps in which the Notch protein is broken down and the tail of the protein acts as a repressor for Notch target genes in the nucleus. The Notch system is known to be involved in a variety of processes, for instance in cell fate specification in the nervous system, where differentiated nerve cells inhibit differentiation of nearby cells, thereby inducing them to become epithelial cells instead [Artavanis-Tsakonas et al., 1999]. Recent work has shown that it could also be responsible for generating stripe-like patterns that give rise to rows of sensory bristles in the *Drosophila* thorax [Corson et al., 2017].

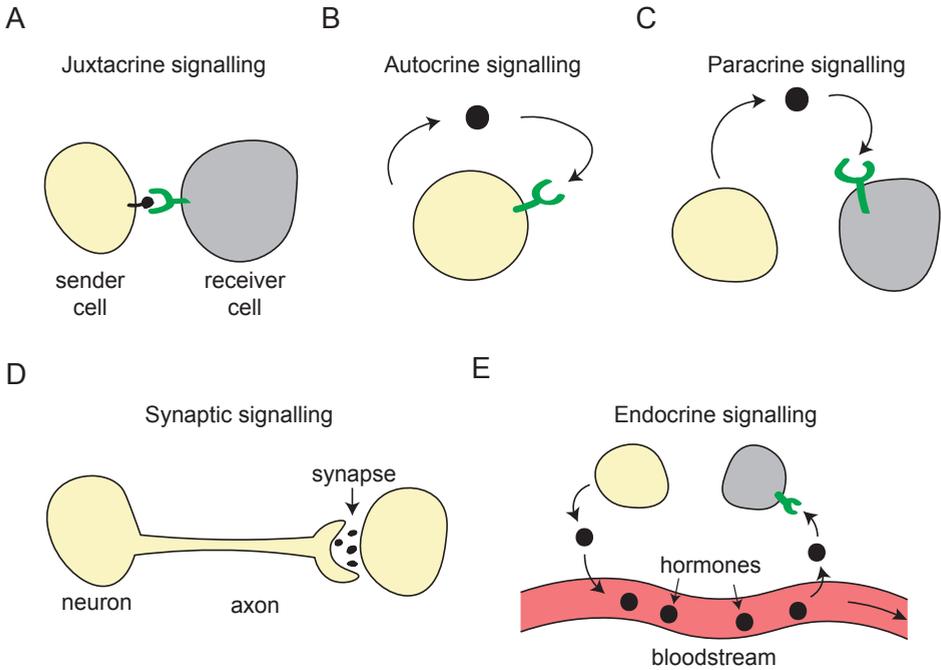
While contact-mediated signaling relies on membrane-bound proteins, most chemical signaling occurs through the secretion of diffusible signaling factors. If the range of diffusion is small, it is possible that the molecules are picked up by the same cell that secreted the molecules. This mechanism of self-communication is termed *autocrine signaling* (Figure 2.1B). One example of autocrine signaling is found in the regulation of proliferation of epithelial cells through epidermal growth factors (EGF) [Sporn & Todaro, 1980]. At low secretion rate of EGF, cells stimulate their own proliferation, whereas overproduction of EGF leads to uncontrollable tumor growth.

In other cases of chemical signaling, signal molecules reach and induce responses in

other nearby cells — we call this type of communication *paracrine signaling* (Figure 2.1C). An interesting example of paracrine signaling is found in the regeneration of hair follicle cells [Chen et al., 2015]. Upon removal of hairs, the hair follicle cells beneath the skin secrete the small signaling molecule Ccl2. This is sensed by macrophages that move towards the affected follicles and secrete Tnf- $\alpha$ , which promotes follicle regeneration. However, due to positive feedback loops, regeneration happens only if sufficiently many hair follicle cells are affected, i.e. when many hairs are plucked at the same time. More generally, autocrine and paracrine signaling are typical in *quorum sensing* systems, in which cells infer cell densities based on the concentrations of secreted molecules in the environment. This may illicit drastic responses as the population density exceed certain threshold [Youk & Lim, 2014a]. A famous example is the sudden transition to bioluminescence in populations of photoluminescent bacteria, where the bacteria suddenly start producing light-emitting luciferase proteins as their population exceeds a threshold [Nealson et al., 1970; Papenfort & Bassler, 2016].

Finally, cells can communicate across distances orders of magnitude larger than the scale of the cells through different mechanisms. Two main examples of such long-distance signaling are *synaptic signaling* in the nervous system (Figure 2.1D), where neurons can pass signals at a speed of 100 m/s, and *endocrine signaling* (Figure 2.1E), which occurs through the transport of secreted hormones through the bloodstream and can occur at a speed of 30 mm/s. It is interesting to note that mechanisms for excitable wave propagation in neurons have similarities to excitation waves in developmental systems of a much smaller scale. Both systems generate waves propagating at far greater speeds than is feasible by diffusion only, and it has been suggested that these different systems could have similar mechanisms for wave generation and propagation [Gelens et al., 2014]. These different modes of signaling are realized by a large variety of different signaling molecules and signal transduction pathways. This confers diversity and specificity to multicellular systems — different systems can have unique signaling machineries and within the same system different ways of signaling can happen simultaneously. Specific molecules may target only specific cells, instead of inducing responses in all cells. Different sized molecules diffuse at different speeds, and may or may not pass through the plasma membrane. The smallest molecules that diffuse across the membrane can directly regulate gene expression by diffusing to the nucleus of the cell. More commonly, however, signaling molecules are sensed by receptors on the cell membrane, which induce signaling transduction pathways that eventually lead to genetic regulation. These pathways may amplify, modulate, integrate or spread different signals. They often involve complicated feedback loops where the components downstream of the receptor regulate each other in various ways [Thurley et al., 2018]. Nevertheless, the net response

of a cell to a sensed concentration is often similar across different systems — the relation between sensed concentration and response can often be well-approximated by a sigmoidal function [Hill, 1910; Ferrell & Ha, 2014a-c]. We will discuss this point in more detail later, as it forms a key assumption of our model for communicating cells (Chapter 3.1).



**Figure 2.1:** Different modes of chemical signaling. Figures based on corresponding cartoons in Alberts et al., 2007. (A) In juxtacrine signaling, cells in direct contact with each other exchange signals. (B) In autocrine signaling, cells secrete molecules that are sensed by themselves. (C) In paracrine signaling, cells secrete molecules that are sensed by other cells. (D) Synaptic signaling is mediated by chemicals that diffuse across synapses between neurons. (E) Endocrine signaling relies on deposition of hormones into the bloodstream and can have effects at long distances.

## MECHANICAL SIGNALING

In recent years, it has been recognized that mechanical interactions play a crucial role in processes such as collective cell migration and development in general [Mammoto & Ingber, 2010; Eyckmans et al., 2011; Serra-Picamal et al., 2012]. Different processes at both the intracellular level as well as the multicellular level require mechanical forces (Figure 2.2A). Cells contract through forces generated by the network of actin and myosin filaments making up the cytoskeleton, which also enables cell motility. In tissues, cells in di-

rect contact with other cells can communicate by propagating mechanical stress, which may enable the formation of wavefronts [Idema et al., 2013; Idema & Liu, 2014]. Cells also interact with the substrate they are on, potentially by deforming it, which in turn indirectly affects other cells in the system (Figure 2.2B). In particular, it has been established that cells can sense the stiffness of their substrate and respond to it by changing their contractility, motility and rate of spreading [Discher et al., 2005]. There are many ways through which cells can sense forces (*mechanosensing*) [Vogel & Sheetz, 2006]. Forces applied to a cell can cause partial protein unfolding, open up ion channels and stabilize receptor-ligand bonds. Cells can also sense local curvature through the release of proteins and the opening and closing of ion channels. Typically, the sensed signals are then transduced inside the cell through biochemical reactions (*mechanotransduction*). This in turn induces downstream responses ranging from cytoskeletal restructuring and cell motility at a fast time scale to altered gene expression and deformation of the extracellular matrix at a slower time scale (*mechanoresponse*).

### TUNNELING NANOTUBES (TNTs)

Contact-mediated signaling can also occur through gap junctions — channels that connect plasma membranes of two cells at close distance to each other — and plasmodesmata, analogous channels that connect plant cells [Alberts et al., 2007]. In recent years, a new type of close-contact signaling has been discovered, which has spurred a wealth of recent papers and on this topic. Tunneling nanotubes (TNTs) are thin tubes joining the plasma membranes of two cells located close to each other, and enable direct exchange of the cytosolic content of the joined cells (Figure 2.2C).

An early example of TNT-like structures was discovered in the *Drosophila* wing imaginal disc, where elongated, filipodia-like extensions of the plasma membrane were observed [Ramirez-Weber, 1999]. These structures — named cytonemes by their discoverers — allow for direct exchange of signaling factors between the joined cells. The term tunneling nanotube was coined in a paper on similar tubular structures between neuronal rat pheochromocytoma cells (PC12) [Rustom et al., 2004]. In particular, it was shown that these structures allow for direct exchange of organelles and membrane vesicles, thus effectively coupling the cytoplasms of two distant cells. Later studies have shown that TNTs can also transfer viruses, proteins and genetic information through e.g. microRNA and mRNA [Hurtig et al., 2010]. Due to its role in RNA exchange, TNTs have been hypothesized to play a role in RNA interference, a mechanism through which small pieces of RNA inhibit translation of particular genes [Belting & Witttrup, 2008]. TNTs have been found to play a key role in stress response and disease. Direct visualization of transport in TNTs have revealed that heat shock diminishes RNA transfer along TNTs, whereas oxidative stress increases it [Haimovich et al., 2017]. In brain tumor cells, TNTs are between

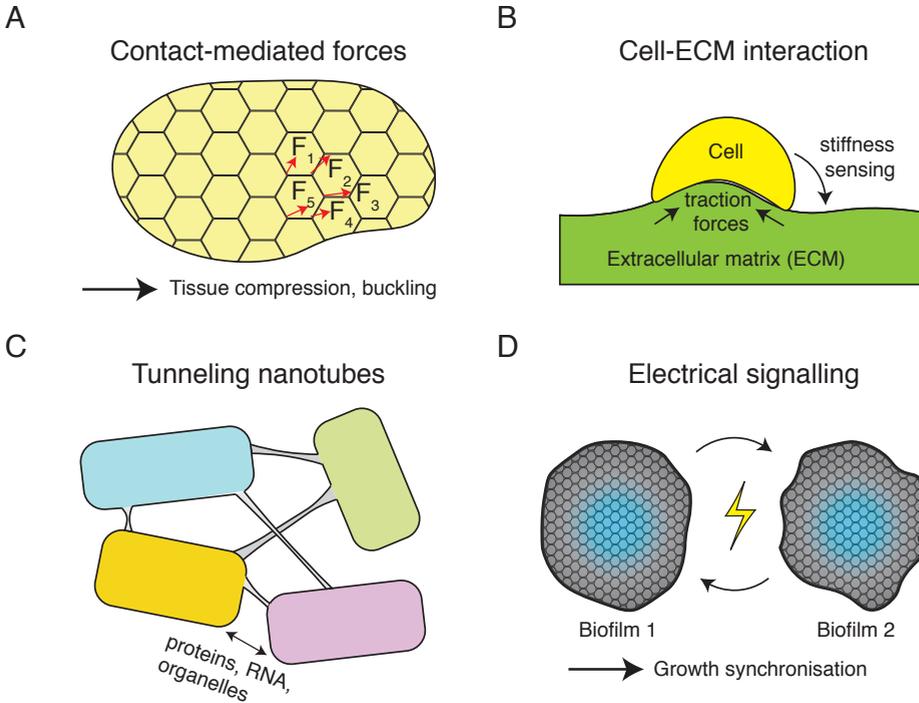
5 and 100 times more abundant than in healthy cells and increase tumor proliferation by making the cells resistant to radiotherapy [Oswald et al., 2015]. Prions — misfolded proteins that are involved in several neurodegenerative diseases — are able to hijack the TNT system to spread across the body, from contaminated food sources to the brain [Gousset et al., 2009]. HIV and other viruses are also known to be capable of spreading through TNTs. These discoveries may have potential applications in the development of new therapeutics targeting these diseases.

In recent years, nanotubes have also been discovered in bacteria, where they seem to play a similar role in exchanging cytoplasmic contents [Baiyda et al., 2018]. They were first discovered in electron microscopy studies of the bacterium *Bacillus subtilis* [Dubey et al., 2011]. Cytosolic GFP molecules produced by engineered strains were observed to spread to wild-type bacteria lacking a gene for its production. In the same model organism, they can allow the spread of antibiotic resistance genes to strains that do not have one. Bacterial nanotubes can also form between bacteria of different species. In one example, *B. subtilis* produces a molecule called WapA which is toxic to the related species *B. megaterium*. By spreading this toxin through nanotubes, *B. subtilis* could inhibit the growth of *B. megaterium* for its own benefit [Stempler et al., 2017]. Thus, bacterial nanotube communication can result in both cooperation and antagonism, depending on the nature of the interaction.

### ELECTROCHEMICAL SIGNALING IN BIOFILMS

Another fascinating example of collective behavior resulting from an unconventional signaling mechanism is found in biofilms of the bacterium *Bacillus subtilis*. In a biofilm, only cells at the periphery have access to nutrients and proliferate quickly, whereas internal cells are the danger of starvation, but have the benefit of being protected from external threats. The survival of the biofilm as a whole thus depends on cooperation between cells in the periphery and at the center. This gives rise to collective oscillations in the biofilm growth dynamics (Figure 2.2D). Phases of slower growth allow the interior cells to feed, whereas in the faster growth regime the exterior cells deplete nutrients [Liu et al., 2015]. This behavior, which bears resemblance to action potentials in neurons, is mediated by bacterial ion channels [Prindle et al., 2015]. These allow for fast propagation of potassium waves responsible for the coordination of growth across the biofilm. Theoretically, these observations have been successfully modeled using delay-differential equations [Martinez-Corral et al., 2018]. The model reproduces the observed sudden onset of oscillations as the size of the biofilm grows — arising as a Hopf bifurcation in the model — as opposed to a gradually increase in oscillation amplitudes. Interestingly, when two biofilms are brought together to grow next to each other, they coordinate their growth dynamics over extended distances through the same electrochemical signaling

mechanism [Liu et al., 2017]. In the presence of high nutrient concentrations, the two biofilms synchronize their growth by oscillating in phase. However, when nutrient availability becomes a limiting factor, competition drives the biofilms to oscillating out of phase, so that each biofilms grows when the other is stationary.



**Figure 2.2:** Non-chemical pattern formation. (A-B) Mechanical forces play an important role in tissue development [Eyckmans et al., 2011]. (A) Cells packed together in a tissue exert mechanical forces to each other, which may induce downstream responses at the gene expression level. (B) Mechanical effects are also mediated through the ECM, whereby cells sense substrate stiffness and actively shape the ECM by applying traction forces. (C) Tunneling nanotubes allow the exchange of proteins, nucleic acids and even whole organelles [Baiya et al., 2018]. (D) Electrical signaling between close-by biofilms leads to synchronization of their growth dynamics [Liu et al., 2017].

## 2.2. SELF-ORGANIZATION IN MULTICELLULAR SYSTEMS

There are myriad examples of multicellular systems in which interactions between genetically identical cells leads to phenotypic heterogeneity (i.e. differences in gene expression level), often through the emergence of self-organized patterns and other collective phenomena. Here, we will discuss examples of self-organizing multicellular systems with an emphasis on two phenomena: aggregation of the slime mold *Dictyostelium discoideum* and the formation of somites in the development of bilateral animals. While

these systems greatly differ in many aspects, the underlying structure and the patterns they generate have a few interesting similarities. Both systems generate synchronized oscillations and traveling waves, which have been modeled as excitable systems using the FitzHugh-Nagumo model [FitzHugh, 1961; Nagumo et al., 1962; Gelens et al., 2014].

### SLIME MOLD AGGREGATION

A classic example of multicellular self-organization is found in the slime mold *Dictyostelium discoideum* (Figure 1.1C). This amoeba switches between an individual and a multicellular state depending on its environment. When nutrients are plentiful, *Dictyostelium* cells maintain individual identities. However, when nutrient becomes limiting, the cells aggregate and form fruiting bodies that look similar to mushrooms. Aggregation is an intricate process in which certain cell centers attract other cells toward them and grow over time. The cells attracted to the centers tend to form spiral wave patterns. This process is mediated by the secretion of a small molecule called cyclic adenosine monophosphate (cAMP). When the cells are starved, they start producing and sending out cAMP molecules. The molecules are released in pulses in an oscillatory manner, and through the extracellular environment the oscillations of different cells become synchronized over time. This behavior has been reproduced in a number of models, mostly based on reaction-diffusion processes [Martiel & Goldbeter, 1987; Palsson & Cox, 1996], but also in individual-based models [Dallon & Othmer, 1997]. In particular, these models and later experimental work revealed an important role for noise in the system to stochastically drive cells to start oscillating.

Recent work on *Dictyostelium* aggregation has revealed a number of interesting results that were previously unknown. Until recently it was not clear whether the oscillations can arise on an individual level or are a purely collective phenomenon. Gregor et al. tested this by subjecting individual cells to varying levels of cAMP, and observed oscillations at sufficiently high concentrations of cAMP [Gregor et al., 2010]. This proved that modulating cAMP levels alone is sufficient to induce the oscillations, and that the system does not rely on more complicated collective mechanisms. Furthermore, the signaling network of *Dictyostelium* has been worked out in more and more detail and is now thought to be largely well-characterized at a molecular level. However, detailed models including all components of these networks have not yielded better results and in fact fail to reproduce certain experimental features. In contrast, a recently introduced model for this process reproduced features in good agreements with experiments [Sgro et al., 2015; Noorbakhsh et al., 2015]. Cells are modeled as coupled units that show an excitation when the cAMP concentration is above a certain threshold. In their model, each cell is modeled as an excitable unit similar to a neuron. The cell detects cAMP in terms of fold-changes, and feeds the signal into an excitable circuit modeled by the FitzHugh-

Nagumo model. Upon spiking, the cell releases a bout of cAMP to its environment which can be sensed by other cells. This model faithfully reproduces the collective oscillations and spiral waves, as well as other experimentally observed features such as oscillation times and responses to single cAMP pulses.

## SOMITE FORMATION

Self-organized patterns are abundant and indispensable in embryonic development. A particularly noteworthy process is body segmentation, which occurs in most animals including arthropods (which includes the insects) and vertebrates (for instance in the formation of the spinal cord). The former has been well-studied in *Drosophila*, where the key genes and regulatory processes behind segmentation have been characterized. The latter has seen a recent surge in interest, especially when it comes to somite formation or somitogenesis. During development in bilateral animals, part of the mesoderm called the presomitic mesoderm (PSM) gives rise to blocks of connected tissue called somites, which in vertebrates eventually form the vertebrae, rib cage and other parts of the back.

The formation of these somites has been studied intensively in the mouse and zebrafish model organisms (for recent reviews, see [Pourquié, 2011; Oates et al., 2012; Hubaud & Pourquié, 2014]). The PSM forms an elongated, U-shaped structure in the anterior-posterior direction, around the neural tube — the precursor of the nervous system. As the embryo develops, the PSM grows in the posterior direction, while somites form one by one at the anterior end. Somitogenesis is accompanied by oscillations of genes involved in various signaling pathways, together with an arresting wavefront that travels from anterior to posterior end. See also Figure 5.5A and the accompanying discussion in Chapter 5.

Numerous theoretical models have attempted to explain various aspects of somitogenesis (for reviews, see [Baker et al., 2008; Tomka et al., 2018]). An early model named the clock-and-wavefront model attempted to explain this phenomenon as the combined result of a synchronous oscillator and a wavefront [Cooke & Zeeman, 1976]. In this model, a “segmentation clock” generates synchronized oscillations among all cells of the PSM. An arresting wavefront then sweeps across the PSM and halts the oscillations. A specific phase (e.g. the peak of the oscillation) could then be responsible for setting the boundaries of the somites.

A number of recent observations have refined this “clock-and-wavefront” picture. First, experiments show a traveling wave of gene expression moving in the opposite direction of the arresting wavefront, which slows down as it approaches the wavefront and comes to a complete halt before reaching it. This wave is distinct from the arresting wavefront in the clock-and-wavefront model. It is not due to cell motion or other active processes,

but arises due to spatially-varying oscillation frequencies leading to phase differences between cells at different locations in the PSM. Furthermore, the collective oscillations arise through a self-organizing process, as was demonstrated in *ex vivo* experiments in which the dissected tail of the PSM, mixed with other cells, was observed to form oscillations and waves again [Tsiairis & Auhlehla, 2016; Hubaud et al., 2017].

How do the oscillations arise in the first place, even at the level of individual cells? It turns out that a delayed negative feedback loop is at work here [Tyson & Novak, 2010; Geva-Zatorsky et al., 2006]. In zebrafish, the negative feedback loop is implemented by the *Her/Hes* system. The *Her/Hes* protein can dimerize and repress its own production, thereby generating sustained oscillations. In mice, negative feedback loops have been identified in three key signaling pathways: FGF, Wnt/ $\beta$ -catenin and Notch. Interestingly, oscillations in these pathways are all linked together, for instance through a shared regulatory element [Hubaud & Pourquié, 2014] or possibly a different mechanism [Sonnen et al., 2017]. Sonnen et al. showed that applying periodic pulses of regulatory molecules of the Wnt pathway also induced oscillations in the Notch pathway, and vice versa. Moreover, oscillations of either pathway become synchronized over time, indicating at some entrainment mechanism which is yet unknown.

How do the oscillations in the PSM synchronize? It has been recognized that Notch signaling plays a key role in maintaining oscillation synchrony. However, it is now thought that the synchrony in somite oscillations is initiated independent of the Notch system, and that Notch signaling is merely responsible for maintaining the synchronized oscillations [Riedel-Kruse et al., 2007]. Moreover, Hubaud et al. recently proposed that the segmentation clock is an excitable system [Hubaud et al., 2017]. They showed that dissociated and reaggregated cells could re-initiate collective oscillations after oscillation arrest. By triggering the onset of oscillations through removal of a Notch inhibitor, they showed that the oscillations resembled those found in excitable systems, showing features like a switch-like activation response and a refractory period before reactivation is possible. What makes the PSM oscillations stop and what is responsible for the traveling wave of gene expression is not yet entirely clear, although different suggestions have been put forward. It is thought that gradients of Wnt, FGF and retinoic acid play a role in arresting the oscillations at the anterior end, in addition to generating an oscillation frequency profile responsible for the traveling waves. However, how these spatial gradients precisely interact with each other and with genetic components of the system is not yet fully understood [Hubaud & Pourquié, 2014].

## OTHER PATTERNS IN DEVELOPMENT

In recent years, quantitative approaches have become more prevalent in developmental biology (reviewed in [Oates et al., 2010; Reeves et al, 2006; Tomlin & Axelrod, 2006;

Krupinski et al., 2012]). Recent studies have shown how a combination of experimental and theoretical work can lead to successful descriptions of developmental patterning. Here we mention a few examples. The formation of the eye disc in *Drosophila* is accompanied by a wavefront — termed the morphogenetic furrow — that sweeps across the system. The locations of the photoreceptor cells are set by the wave to form a regularly shaped crystal lattice. Many of the signals involved in the generation and propagation of the morphogenetic furrow and their mutual interactions have been found, leading to the development of theoretical models that capture this essence of this dynamics [Lubensky et al., 2011, Fried et al, 2016]. Similarly, traveling waves are found in the brain development of *Drosophila*, where they regulate the differentiation of neural stem cells into neuroepithelium cells. This behavior has been successfully described in a theoretical model involving a simple regulatory circuit together with juxtacrine Notch signaling, also through an excitable mechanism [Jörg et al., 2019]. A neat example of theoretical modeling to explain pattern formation is found in the example of *Drosophila* bristle formation — this has been shown to arise from a local activation, lateral inhibition mechanism involving the Notch signaling pathway [Corson et al., 2017]. Finally, the early development of *Drosophila* requires traveling waves mediated by mechanical interactions — this example is discussed in the following paragraph.

### MECHANICALLY-INDUCED PATTERNS

Mechanical interactions can also give rise to dynamic patterns such as mechanical and chemical waves. The first way in which mechanics contributes to patterning is through forces between cells or within cells (e.g. through the actomyosin cytoskeleton). For example, in an expanding group of epithelium cells, it has been shown that stress between cells can propagate as a mechanical wave through the tissue [Serra-Picamal et al., 2012]. Furthermore, in the early *Drosophila* embryo, nuclear divisions take place without cytokinesis. In other words, the embryo develops into a large single cell with many nuclei before these separate to become a multicellular embryo — the multinucleated cell is called a syncytial embryo. To coordinate cell divisions across a very large cell, the cell generates mitotic trigger waves which travel across the syncytial embryo and synchronize rounds of cell division [Chang & Ferrell, 2013]. A theoretical model which treats the embryo as an excitable medium shows that the waves could emerge from purely mechanical interactions [Idema et al, 2013].

The second way in which mechanics enters is through cell sorting mechanisms. When cells of different types are isolated from an organism and mixed together in vitro, the initially mixed configuration of cells will tend to unmix into separate layers constituting cells of the same type. This process, which resembles what happens if one mixes together immiscible liquids, has been hypothesized to rely on a thermodynamic process

which minimizes a free energy that depends on the surface and interfacial tensions of the cells. This idea is called the “differential adhesion hypothesis” (DAH) (reviewed in [Steinberg, 2007; Amack & Manning, 2012]). In recent years, adhesion between cells has been related to cadherin molecules, which allow cells to stick to each other. However, taking into account only cadherin molecules alone appeared to be insufficient to explain the total adhesion energy between cells. This has prompted a revised DAH, which proposed that differences between cells at the edge of the tissue with the rest of the tissue, or “mechanical polarization” as the authors call it, plays an important role in this process [Amack & Manning, 2012].

This idea of sorting cells based on their adhesive properties has also given rise to engineered multicellular systems with programmable pattern properties [Toda et al., 2018]. Here, the researchers built a modular synthetic system where juxtacrine cell-cell signaling could induce differences in not only gene expression and cell type, but also in the adhesive properties of the cell. By tuning the synthetic circuit underlying the signaling system, they were capable of generating a diverse range of self-organized patterns formed by cells sorting themselves in space due to differences in adhesive properties.

#### SELF-ORGANIZATION IN BACTERIA

Bacteria have long been known to respond to externally added chemical cues by generating intricate patterns [Budrene & Berg, 1991]. For instance, chemotaxis can drive bacteria to form patterns when nutrients are heterogeneously distributed in space. However, bacterial systems where collective behavior arises through cell-cell communication have been studied only relatively recently. Quorum sensing in bacteria was discovered in the marine bacterium *Aliivibrio fischeri*. In this system, the bacteria produce luciferase — a bioluminescent protein — only if their density exceeds a certain threshold. These studies have led to the identification of genetic networks and molecular details of bacterial quorum sensing systems [Waters & Bassler, 2005]. This in turn enabled the development of synthetic quorum sensing systems, for instance by engineering quorum sensing machinery into bacterial species that naturally do not have such a system. By manipulating their network architecture, collective behavior such as synchronized oscillations and traveling waves could be engineered in a strain of *E. coli* [Danino et al., 2010].

#### FURTHER READING

A number of recent articles review multicellular pattern formation from different perspectives. [Schweisguth & Corson, 2019] discuss pattern formation in various developmental systems from both a theoretical and an experimental point of view. Pattern formation in microorganisms such as bacterial and social amoeba has a long history.

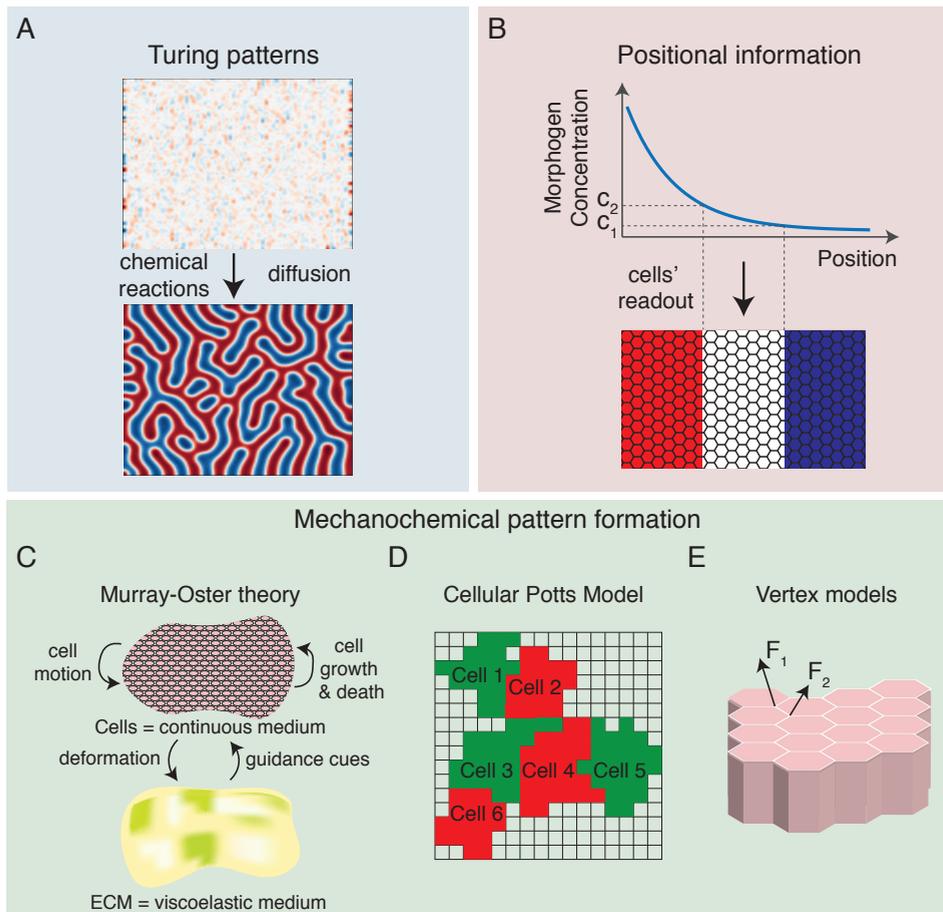
The lengthy review of [Ben-Jacob et al., 2000] covers both experimental observations and theoretical modeling using mostly reaction-diffusion equations. [Von Bronk et al., 2018] also review interacting microbial systems, but mainly from an ecological perspective and in the context of biofilms. [Howard et al., 2011] discuss how active processes generate patterns through force-generating molecular motors and the coupling between biochemical and mechanical processes. [Lander, 2011] discusses how systems biology concepts and engineering approaches can be applied to understand pattern formation and growth in development. Synthetic approaches to engineer patterns in biological systems are the topic of several reviews, including [Teague et al., 2016; Scholes & Isalan, 2017; Santos-Moreno & Schaerli, 2018]. In the field of tissue engineering, pattern formation in stem cells is a hot topic [Heemskerk & Warmflash, 2016], especially in the context of making artificial organ-like structures called organoids [Sasai, 2013; Lancaster & Knoblich, 2017].

## 2.3. MODELS AND MECHANISMS OF MULTICELLULAR PATTERN FORMATION

In this section, we review the main theoretical frameworks that have been proposed to explain pattern formation in multicellular systems. Two early theories — the Turing mechanism and positional information — have traditionally been regarded as alternative explanations for pattern formation. However, it has recently been recognized that they likely both play an important role in pattern formation [Green & Sharpe, 2015]. The theories also share an important role for chemical signals (morphogens), either as self-organized patterns or through gradients induced by external effects. As discussed earlier, mechanical interactions between cells and between cells and the extracellular matrix can also drive self-organized structures to form. Hence, in complex multicellular systems one should expect different mechanisms to be at work and to interact with each other [Schweisguth & Corson, 2019]. In development, pattern formation is followed by morphogenesis — the embryo's acquisition of a shape. While chemical cues are mainly responsible for the former, the latter requires carefully orchestrated motile processes — such as collective cell migration — which rely on mechanical interactions. These observations have spurred the development of hybrid theoretical models that combine mechanical and chemical interactions [Recho et al., 2019].

### REACTION-DIFFUSION THEORY

The classical and dominant idea about pattern formation is based on a mathematical framework developed by Alan Turing, published one year before the discovery of the DNA double helix [Turing, 1952] (Figure 2.3A). The core argument of Turing's analysis



**Figure 2.3:** Mechanisms of pattern formation. (A) Reaction-diffusion systems consisting of chemicals that interact and diffuse self-organize into regular patterns through the Turing mechanism. Plots obtained using code from the GitHub repository TuringPatterns. (B) Morphogen gradients confer positional information to cells and induce downstream gene expression changes, which may lead to discrete states arising in a cell population (forming the colors of the French flag in the picture) [Wolpert, 1971]. (C-E) Models of mechanochemical pattern formation, whereby mechanical and chemical interactions are linked and together give rise to patterns. (C) The classical Murray-Oster theory provides a continuum description of cell and ECM densities, together with various processes coupling and adjusting these quantities [Murray, 2003]. (D) The Cellular Potts Model describes how cells change their shapes over time by minimizing an energy function [Merks & Glazier, 2005]. (E) Vertex models describe how tissue shapes change by calculating forces at each vertex shared by multiple cells [Alt et al., 2017].

is based on a clever mathematical analysis of a system of chemicals that diffuse freely and may react with each other in processes described by reaction rates. Such systems are generally known as reaction-diffusion systems and are described by coupled partial differential equations (PDEs) for each of the chemical species under consideration. Un-

der general assumptions, Turing decomposed the spatial information about the concentrations of these chemicals into Fourier modes, and showed that some of these modes could grow whereas others decayed. If at least one mode grows, then the system will eventually form a spatial pattern at the wavelength of the fastest growing mode — this phenomenon is referred to as the Turing instability. The resulting patterns can look like spots or stripes, depending on the details of the system and the initial conditions. The idea is that the patterns of these chemicals — named morphogens by Turing — could induce changes in gene expression, which in turn could lead to morphogenetic processes and cell differentiation.

The power and appeal of this framework lies in the fact that the conditions under which patterns can form can be reduced into a set of simple mathematical statements in terms of the parameters of the system. Later, Gierer and Meinhardt showed that Turing patterns arise in a simple two-chemical system, where one chemical is a short-range activator and the other is a long-range repressor [Gierer & Meinhardt, 1972]. In this model, the condition Turing derived simplifies to the statement that the inhibitor should diffuse much further than the activator.

Experimental studies have reproduced the patterns predicted by Turing's framework in systems such as chemical reactors [Castets et al., 1990], which are easier to control due to the absence of biological material. Theoretical studies of reaction-diffusion have produced patterns that are strikingly similar to patterns observed in biological systems such as on the coats of animals [Murray, 2003]. This is especially striking when studied on specific geometries that closely mimic the shape of the biological system, such as sea shells and tails of animals, or on snakes [Murray & Myerscough, 1991]. Nevertheless, the idea that Turing patterns arise in biology remained purely theoretical for a long time, because the molecular identities of the diffusing molecules that could create these patterns have remained elusive for a long time. Furthermore, Turing patterns are typically not robust with respect to perturbations and changes in parameters, as will be discussed later. This has led to scepticism about the usefulness of the Turing framework in explaining biological patterns, even from the most fervent practitioners of reaction-diffusion modeling [Maini et al., 2012].

Progress in recent years has led to the identification of a number of potential Turing systems, where the molecular identities and interactions of a short-range activator and long-range inhibitor have been discovered (for recent reviews, see [Kondo & Miura, 2010; Marcon & Sharpe, 2012]). Notable examples include hair follicle positioning [Sick et al., 2006], digit patterning [Sheth et al., 2012, Raspopovic et al., 2014], stripe formation in the mammalian palate [Economou et al., 2011] and zebrafish skin patterns [Watanabe & Kondo, 2015; Mahalwar et al., 2018]. In general, these studies largely rely on performing experimental perturbations, such as deletion or overexpression of certain genes,

which show that the self-organized patterns respond to these perturbations in a way that matches simulations of Turing models. However, these studies have also revealed that most patterning processes are likely more complex than simple activator-inhibitor systems of two chemicals. For instance, in the case of zebrafish skin patterns, there are at least three different types of pigment cells. They interact by promoting or repressing the migration or growth of cells of the same or a different type. Hence these interactions are also not purely chemical, yet the patterns they generate are well-recapitulated in a reaction-diffusion model of the system [Wantanabe & Kondo, 2015]. Furthermore, the patterns may rely on more than two diffusing molecules, as is in the case of digit patterning, where a three-molecule gene circuit has been discovered [Raspopovic et al., 2014]. However, note that this system still contains a core motif with a short-range activator and long-range inhibitor. As such, we can view these systems as the first plausible examples of biological Turing patterns that have some experimental support, even though their patterning processes are more complex than described in Turing's original model.

#### LIMITATIONS OF REACTION-DIFFUSION MODELS

Despite these recent successes in identifying potential biological Turing systems, using reaction-diffusion theory to explain biological pattern formation has a number of drawbacks and limitations. First, reaction-diffusion models model pattern formation as a purely chemical process without any direct reference to relevant biological processes. While it can be argued that it is implicitly assumed that the molecules are secreted by cells and the molecular interactions may be coupled to genetic interactions, these features are not explicitly modeled and parameters controlling them cannot be adjusted. This description would be perfectly accurate when signaling molecules can freely diffuse inside a tissue, and react instantaneously with other molecules both within and outside of cells. But in biological systems, these molecules encounter cell membranes, bind to receptors, induce signal transduction pathways and may eventually regulate gene expression. These essential processes are neglected altogether in the reaction-diffusion models, which assume that cells passively respond to their chemical environment without taking any active role in shape it (at least, this is not modeled explicitly).

Part of this limitation stems from the fact that reaction-diffusion systems assume a continuous field of cells rather than discrete cells. This makes it difficult to take into account processes that play a role inside individual cells. Furthermore, such continuum descriptions cannot properly take into account cell-to-cell heterogeneity, which plays a role in any multicellular system, even when the cells are genetically identical [Ackermann, 2015]. One alternative is then to construct individual-based models (IBMs; also called cell-based or agent-based models) where cells are modeled as individual units

that interact with each other [Deutsch & Dormann, 2005]. A particular type of IBM is a cellular automaton, which is a discrete model where local interaction rules specify how cells change over time. Specifically, cellular automata model cells with discrete states on a discrete space such as a lattice, that evolve according to discrete time steps. Until recently, IBMs for patterning have often assumed highly simplified rules for how the agent-based evolves, which are typically derived from phenomenological considerations [Wolfram, 1984]. However, recent work on pattern formation on the lizard skin shows that combining cellular automata with reaction-diffusion processes can successfully reproduce observed patterns [Manukyan et al., 2017]. This suggests that extending IBMs such as cellular automata to include more realistic features of biological systems could provide rich ground for further investigation.

The second major issue of the Turing framework is that Turing patterns tend to suffer from a lack of robustness. In this context, it means that only specific, fine-tuned parameters lead to pattern formation. Hence, under a mild perturbation that changes some parameters, it would not be unlikely that the entire pattern dissolves. It has been known for a long time that typically only a small portion of the parameter space is capable of generating Turing patterns [Murray, 1982]. Extending reaction-diffusion models to include effects of growing domains, stochastic effects and time delays does not resolve this problem [Maini et al., 2012].

These early findings were done for specific, small systems (e.g. activator-inhibitor models). In contrast, recent high-throughput numerical studies have looked for Turing patterns in larger networks consisting of more interacting molecules [Cotterell & Sharpe, 2010; Marcon et al., 2016; Zheng et al., 2016; Scholes et al., 2019]. These studies constructed libraries of all possible networks up to a certain number of diffusing molecules and quantified each network's capacity to form patterns. This was done by simulating each network with a large set of random parameters and initial conditions, and by obtaining statistics on these simulations. In particular, one study identified a set of six distinct mechanisms — distinguished by their core network motifs — that underlie most of the networks capable of generating Turing patterns [Cotterell & Sharpe, 2010]. It has also been shown that differential diffusivity — e.g. having a fast diffusing activator and slowly diffusing inhibitor — is not required, but that Turing patterns can also arise if both molecules diffuse equally far [Marcon et al., 2016]. Furthermore, certain types of networks with three interacting signaling molecules have higher robustness than those with only two [Marcon et al., 2016]. However, the model set up in this study and thus also its claimed findings were criticized in a later study, which performed a similar network enumeration with a more general model, but did not find any significant increase in robustness as the number of molecules increased [Scholes et al., 2019]. Altogether, these results show that even in Turing models, the exact conditions (in terms of gene

networks) under which patterns form are not fully understood.

## POSITIONAL INFORMATION

2

The other classical theory about pattern formation, proposed by Lewis Wolpert, is centered around a concept called positional information [Wolpert, 1969]. The main idea is that morphogens can form gradients, so cells at different locations sense different concentrations, which in turn induces differences in gene expression. As such, the morphogen concentration readout of each cell contains information about its position. If we assume that cells have three gene expression states and that their sensed morphogen concentration determines which of the three states they acquire, then we would obtain a French flag pattern (Figure 2.3B).

A classic process in which positional information has been invoked to explain development is the segmentation of the *Drosophila* embryo. Specifically, the segmentation of the body along the anterior-posterior (AP) axes is orchestrated by a set of patterning genes whose interactions are well-characterized. The process starts with the establishment of morphogen gradients through interaction of the embryo with the maternal environment — an example is the *bicoid* gradient along the AP axis [Driever & Nüsslein-Volhard, 1989]. The gradients provide positional information by informing each cell of its position along the AP axis. They induce downstream gap genes, each of which is expressed in largely non-overlapping regions along the AP axis. Further downstream, the gap genes turn on genes specifying the exact positioning of the segments.

In recent years, technical advances have made it possible to accurately measure protein and mRNA concentrations *in vivo* up to the single-molecule level [Raj et al., 2008]. Recent studies relying on these and other new techniques have studied the diffusion of *bicoid* proteins *in vivo* [Gregor et al., 2005], the stability of the *bicoid* gradient [Gregor et al., 2007a] and the precision of the gradient [Gregor et al., 2007b]. One particular conclusion from these studies is that the *bicoid* gradient is surprisingly precise, allowing the cells to sense concentrations with accuracies close to their physical limits. This allows each row of cells along the AP axes to assume a separate identity over time. In this context, information theory provides an interpretation of this result that has biological significance [Tkačik & Walczak, 2011]. Using the concept of mutual information, the authors quantified the amount of information that the *bicoid* concentration provides to one of the downstream gap genes. This calculation showed that with four gap genes each individually reading out the gradient, the system could provide sufficient positional information for each row of cells to ‘know’ their precise location along the AP axis.

While positional information is often displayed as an alternative to the Turing mechanism, it should be recognized that in many biological systems, both mechanisms are at play [Green & Sharpe, 2015]. For example, in the development of digits such as fingers

in the mouse, it has been proposed that the digit pattern arises from a reaction-diffusion process, but that an FGF gradient sets the spatially-varying wavelength of this pattern.

### MECHANOCHEMICAL PATTERN FORMATION

In his seminal 1952 paper, Turing recognized the importance of mechanical forces in shaping the development of an embryo, but omitted a description of the mechanical interactions because “The interdependence of the chemical and mechanical data adds enormously to the difficulty” [Turing, 1952]. His remark has not remained unnoticed, and in the course of the following decades progress has been made in developing theoretical models that describe mechanical interactions at different levels of detail [Wyczalkowski et al., 2012]. This has been accompanied by tremendous progress in our experimental capabilities in precisely measuring forces between and inside cells [Serrapicamal et al., 2012].

An early attempt to incorporate mechanical forces into a pattern formation theory is the continuum mechanics approach known as the *Murray-Oster theory* ([Oster et al., 1983]; reviewed in [Murray, 2003]; also see Figure 2.3C). In their model, the cells and the extracellular medium (ECM) are modeled as continuous fields subject to various forces and interactions. Cells migrate on the ECM as a result of various processes, including convection (passive migration due to deformation of the ECM), random dispersal, haptotaxis (moving up adhesiveness gradients) and chemotaxis (moving up chemical gradients). The interaction between cells and the medium is described by a mechanical equilibrium equation that takes the viscosity and elasticity of the ECM into account, as well as cell traction forces (applied by the cell to the ECM) and other forces. Finally, the total ECM material is assumed to be conserved. These elements are brought together in a set of coupled PDEs, which can then be solved numerically. The Murray-Oster theory assumes that the ECM deformations are small by modeling the ECM as a linear viscoelastic medium, but this is not necessarily true in general. Later models have gone beyond this assumption and developed full non-linear theories [Rodriguez et al., 1994]. The Murray-Oster theory and its extensions have been applied to modeling the formation of periodic patterns of feather germs, fingerprint formation, the formation of microvilli, gastrulation (in particular, ventral furrow formation), brain cortical folding and many other phenomena [Murray, 2003; Wyczalkowski et al., 2012].

As a continuum mechanics theory, the Murray-Oster theory also does not take into account individual cells of the system. The alternative is to develop cell-based models that do model interactions between individual cells. A particularly powerful cell-based model is the *Cellular Potts Model*, originally introduced by Graner and Glazier [Graner & Glazier, 1992] (Figure 2.3D). This model was inspired by the earlier mentioned dif-

ferential adhesion hypothesis and reproduces the experimentally observed cell sorting processes under suitable parameters [Merks & Glazier, 2005]. While many variants of this model exist, its classical description consists of a lattice where each site has a state specifying which cell it belongs to. Lattice sites with the same state together constitute a single cell. At each time step, the shape of one of the cells is changed probabilistically according to the interaction energies arising from surface and interfacial tensions. These energies are captured into a single mathematical function, the Hamiltonian of the system, which is used to determine probabilistically how the system evolves. Effects such as haptotaxis, cell differentiation, cell division and apoptosis can also be incorporated into this framework [Merks & Glazier, 2005]. The Cellular Potts Model has been applied to a wealth of biological processes, including evolution of morphogenesis [Hogeweg, 2000], tumor growth and evolution [Szabó & Merks, 2013] and the formation of vascular networks [Scianna et al., 2012].

Another class of cell-based models focusing on mechanical interactions are *vertex models* (reviewed in [Alt et al., 2017; Fletcher et al., 2014]; Figure 2.3E). In vertex models, cells are modeled as 2D polygons or 3D polyhedral which are in direct contact with each other. For instance, in 2D they could form a hexagonal lattice. The forces between the cells are determined by the changes in volume, surface area, line tension and possibly external forces. The forces at each vertex are added up to calculate how the shape of the entire tissue changes over time. Vertex-based models have been used to model epithelial morphogenesis, cell migration, gastrulation, appendage formation and many other morphogenetic processes.

#### OTHER MODELS OF PATTERN FORMATION

In another class of models, patterns arise from the dynamic evolution of a system under a set of simple rules — these are sometimes referred to as “rule-based” patterns [Scholes & Isalan, 2017]. An early example of this are the L-systems proposed by Lindenmayer (reviewed in [Prusinkiewicz & Lindenmayer, 1996]), which produce branching patterns that strikingly resemble plants such as ferns. More recently, rule-based models have been applied to explain features in branching processes in development (reviewed in [Hannezo & Simons, 2018]). In particular, it has been proposed that different branching processes can be explained by a single unifying framework based on random walk theory [Hannezo et al., 2017].

In the field of active matter, the interactions between active “particles” gives rise to collective phenomena such as collective motion. A famous example is the Vicsek model, an individual-based model with simple interaction rules that reproduces patterns of bird flocking [Vicsek et al., 1995]. The success of this model is partly due to the theo-

retical success in deriving coarse-grained, continuum model from this IBM, casting it into a form similar to the Navier-Stokes equations in hydrodynamics [Toner & Tu, 1998]. At a smaller scale, active processes including advection and active transport have been proposed to play a key role in intracellular pattern formation, for instance in the establishment of morphogen gradients and reorganization of the cytoskeleton [Howard et al., 2011].

It should be noted that different processes can lead to the same type of pattern, meaning that one cannot always infer the underlying process by only examining the final pattern [Hiscock & Megason, 2015]. Activation-inhibition systems producing Turing-like patterns can arise through chemical interactions, mechanical interactions or through cell motility. For instance, traction forces applied by cells to the ECM could be responsible for local activation, whereas the elasticity of the ECM corresponds to long-range inhibition. Alternatively, cell growth in a confined environment produces compressive forces which can lead to periodic buckling of a tissue sheet. This mechanism is at play in the formation of gut villi, which form an undulating pattern with regularly spaced intervals [Shyer et al., 2015]. Thus, to distinguish between different types of patterning mechanisms, one should examine different features of the system, such as response to perturbations and the dynamics of pattern formation [Hiscock & Megason, 2015].

# II

## **COLLECTIVE DYNAMICS OF COMMUNICATING CELLS: MODELING AND ANALYSIS**

# 3

## CELLS COMMUNICATING WITH ONE SIGNALING MOLECULE GENERATE DIVERSE COLLECTIVE BEHAVIORS

*We present a theoretical framework and mathematical analysis of a model of spatially distributed secrete-and-sense cells that communicate through the same signaling molecule. Starting from an earlier defined model (Section 3.1), we first extended the analysis to cells with continuous response functions (instead of binary cells) and analyzed the model using tools from dynamical systems theory (Section 3.2). This revealed a cell-density-dependent transition between an autonomous phase — in which cells act independently from each other — and a collective phase — in which cells synchronize their gene expression. Then we study the effect of local perturbations on the system, by quantifying sensitivity to initial conditions using metrics tailored to the model under study (Section 3.3). Finally, we end this chapter by proposing the concept of mutual information between initial state and final state as a metric to quantify the system's degree of “autonomy”. (Section 3.4).*

### **3.1. MULTI-SCALE MODEL FOR QUORUM SENSING CELLS**

In this section, we introduce a theoretical model for communicating cells that forms the basis of all subsequent work presented in this thesis. This model was first conceived and studied in [Maire & Youk, 2015a], where the authors showed that concepts such as auton-

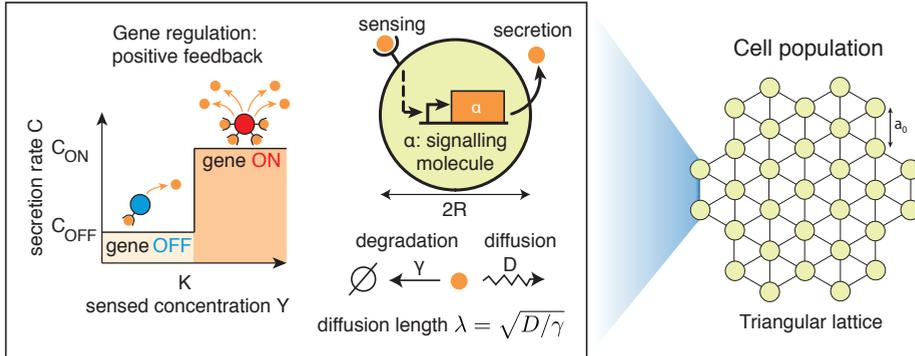
omy and collectiveness can be quantified and tuned by adjusting microscopic features of the system (i.e. at the level of individual cells). Specifically, they introduced the concept of ‘phenotype diagrams’ that summarize the collective response of the population as a function of molecular interaction parameters such as the distance between neighboring cells and characteristics of the signaling circuit. They then define an ‘entropy of population’ that quantifies the number of steady state configurations the system can have, which relates to its ability to form patterns. While this earlier work provides a consistent theoretical framework, the model makes numerous assumptions that limit the scope and extent of the model. Our aim throughout this thesis is thus two-fold: first, we extend the theoretical framework of [Maire & Youk, 2015a] to study more complex interactions and realistic assumptions. Simultaneously, we develop and apply various other techniques to more fully understand emergent phenomena and ensemble-level properties that arise from this model. We also ask what new quantities and concepts are useful for understanding the collective behavior of this complex system.

Consider a system of  $N$  identical cells, which are spherical and have a radius  $R_{\text{cell}}$ . Let the cells be arranged on a hexagonal or triangular lattice with distance  $a_0$  between nearest neighbors (Fig. 3.1). The cells secrete signaling molecules which are sensed both by the secreting cell itself (autocrine signaling) and by other cells in the system (paracrine signaling). The signaling molecules are characterized by a diffusion constant  $D$  and a constant degradation rate  $\gamma$ . We assume that the molecules diffuse freely in the medium between the cells as a first-order approximation, but note that in many biological systems diffusion is anomalous, meaning that the mean-squared displacement of the molecules scales with time either sub-linearly (subdiffusion, e.g. due to macromolecular crowding, [Tolić-Nørrelykke et al.; Golding & Cox, 2006]) or super-linearly (superdiffusion, e.g. for actively driven motion, [Reverey et al., 2015]). Degradation of the signaling molecules occurs either actively (e.g. through proteases) or due to natural decay of the molecules. We assume that the cells continuously secrete signaling molecules, at a rate that is constant during the time the signaling molecule concentration requires to reach equilibrium. It can then be shown that the steady state profile for the signaling molecule concentration surrounding a single cell secreting at rate  $\eta$  is then [Maire & Youk, 2015a]

$$c(r) = c_R f(r), \quad \text{where } c_R = \frac{\eta\gamma}{4\pi R_{\text{cell}}\lambda(\lambda + R)}, \quad \text{and } f(r) = \frac{R_{\text{cell}}}{r} e^{(R_{\text{cell}}-r)/\lambda} \quad (3.1)$$

Here we introduced  $\lambda = \sqrt{D/\gamma}$  as the characteristic diffusion length of the signaling molecule.

Next, suppose that the cells respond to sensed signaling molecules by changing their



**Figure 3.1:** Schematic of the agent-based multicellular model with indicated model parameters. Adapted from [Maire & Youk, 2015a] and [Olimpio, 2016].

secretion rate, thereby creating a feedback in the system that allows the cells to dynamically change their secretion rate. Assume that the cells update their secretion rate on a time scale that is much larger than the time scale on which the signaling molecules reach steady state. This is a realistic assumption in multicellular systems, where morphogen gradients are typically quickly established, but cells take considerably longer time to settle into distinct cell states [Heemskerk et al., 2019].

Furthermore, suppose that the cells respond to sensed concentrations in a step-like fashion, such that the cells secrete at either a high rate  $C_{ON}$  or a low rate  $C_{OFF}$ , determined by whether the sensed concentration is above or below a given threshold  $K$  (Fig. 3.1). If the feedback is positive (activating), then cells will secrete at a high rate if their sensed concentration is above the threshold  $K$ . If the feedback is negative (repressive), then cells will secrete at a low rate if their sensed concentration is above the threshold  $K$ . We first considered digital cells for two reasons. First, experimental studies have shown that signal transduction pathways such as MAPK or other phospho-relay cascades that are triggered by ligand-bound receptors and control gene-expressions downstream, as in our digital cells (Figure 5.1C), can have an effective Hill coefficient with a value of 4 or more (e.g., as high as 32 [Trunnell et al., 2011]). An effective Hill coefficient characterizes the "sharpness" of cell's response to a ligand [Ferrell & Ha, 2014; Plotnikov et al., 2011; Trunnell et al. 2011]. Such high numbers are due to multiple molecular parts amplifying each other's effects in combination. A digital (ON/OFF) response models such high-valued Hill coefficients. The second reason is that a digital response simplifies the mathematics that describes the response, while retaining its main qualitative features,

even when the actual Hill coefficient of the system being modeled is relatively low [Alon, 2007].

In this description, the cells have binary states, which we can identify as ‘OFF’ (if they secrete at the low rate) and ‘ON’ (if they secrete at the high rate). In the following, we measure all distances in units of  $\lambda$  and all concentrations in units of  $C_{OFF}$ . Hence we set  $\lambda = 1$  and set  $c_R = 1$  in Eq. 3.1 for an OFF-cell and  $c_R = C_{ON}$  for an ON-cell. Since distances on the lattice are measured in units of  $a_0$ , we also introduce a dimensionless variable for the cell radius  $r_{\text{cell}} \equiv R_{\text{cell}}/a_0$ . We furthermore take time to be discrete, and assume that the cells simultaneously update their states. Therefore, at any time step  $t$  the state of the system is specified by  $\mathbf{X}(t) = \{X_k(t)\}_{k=1}^N$ , where  $X_k \in \{0, 1\}$  is a binary variable specifying whether cell  $k$  is OFF ( $X_k = 0$ ) or ON ( $X_k = 1$ ). We can then write the secretion rate of the cell as

$$C(X_k) \equiv (C_{ON} - C_{OFF})X_k + C_{OFF}. \quad (3.2)$$

Note that with this definition,  $C(X_k = 1) = C_{ON}$  and  $C(X_k = 0) = C_{OFF} = 1$ . The signaling molecule concentration a cell senses is then the sum of the contributions from all cells in the system, with larger contributions from cells which are close than from cells which are far away. In particular, the concentration a cell  $k$  senses that is due to cell  $m$  can be written as  $f_{km}C(X_k)$ , where

$$f_{km} \equiv \begin{cases} f(r_{km}) & (k \neq m) \\ 1 & (k = m) \end{cases}, \quad (3.3)$$

is an effective interaction between cells  $m$  and  $k$ , which in the case of a single molecule is symmetric (i.e.  $f_{km} = f_{mk}$ ). The total the sensed concentration of a cell  $k$  is then

$$Y_k = \sum_{m=1}^N f_{km}C_m. \quad (3.4)$$

For later reference, introduce an interaction strength  $f_N$  as

$$f_N = \sum_{m \neq k} f_{km}. \quad (3.5)$$

This interaction strength is a measure for how strongly the cells communicate with each other, and is a uniformly decreasing function of  $a_0$ . Note that if all cells secrete at the same rate  $C$ , then they would all sense a concentration

$$Y = (1 + f_N)C. \quad (3.6)$$

Finally, the binary response of the cells can be written as

$$X_k(t+1) = \begin{cases} \theta(Y_k(t) - K) & \text{if interaction is positive} \\ \theta(K - Y_k(t)) & \text{if interaction is negative.} \end{cases} \quad (3.7)$$

We typically continue updating the system until we reach a steady state, where the state of all cells does not change upon updating anymore. Alternatively, the system may settle into a limit cycle, where it cycles through a fixed number of states repeatedly such as in the case of an oscillation. This will be discussed in more detail in the next section.

To summarize, our model is implemented as follows:

1. Initiate the system in some state  $\mathbf{X} = \{X_k\}_{k=1}^N$ .
2. Compute the sensed concentration for each of the cells in the system using Eq. 3.4.
3. Simultaneously update the state of all cells according to Eq. 3.7.
4. If the state of the system has changed, go repeat steps 2-3.
5. Else, terminate the simulation.

### 3.2. SYNCHRONIZATION AND AUTONOMY IN COMMUNICATING ANALOGUE CELLS

Up until this point, we have only considered cells that respond to signaling molecules in a binary fashion, by upregulating or downregulating gene expression levels depending on whether their sensed concentrations are above a certain threshold. This is in fact a rather crude approximation based on the observation that many biological processes have binary outcomes: cells fate determination, apoptosis and allosteric switching are examples of processes that typically have an all or none outcome. However, rather than requiring a strictly binary response curve, bistability in fact only requires sigmoidal rather than hyperbolic (Michaelis-Menten) response functions, a phenomenon which is known as ultrasensitivity [Ferrell & Ha, 2014a]. In recent years, the biochemical basis of ultrasensitivity has been uncovered in a variety of biological systems. Several distinct mechanisms can induce ultrasensitivity, including multisite phosphorylation, stoichiometric inhibitors, and positive feedback [Ferrell & Ha, 2014b]. The response function describes the system responds to varying level of a stimulus. An example is the phosphorylation of a signaling molecule  $X$  through kinase. In this case, the response  $R$  is the concentration of phosphorlylated  $X$  and the stimulus  $S$  is the concentration of kinase. Mathematically, such response functions are typically modeled by the Hill equa-

tion:

$$R = \frac{S^n}{K^n + S^n}. \quad (3.8)$$

Here,  $n$  is the Hill coefficient of the system and  $K$  is a phenomenological parameter that gives the stimulus concentration at which the response is at half of the maximum value. Note that unlike in cooperative binding processes, where the Hill equation can be derived from the underlying reaction kinetics, the Hill equation in 3.8 is a phenomenological equation based on experimental observations of signal response curves. The Hill coefficient controls the steepness of the sigmoidal response — higher values correspond to steeper response curves exhibiting higher degrees of ultrasensitivity. The value of the Hill coefficient varies drastically between different ultrasensitive signaling transduction pathways and can be as low as 1.3 and as high as 31 [Ferrell & Ha, 2014a; Trunnell et al., 2011].

These findings call for a more elaborate study of the dynamics of our system with a sigmoidal function rather than step function as response function. This extension was first considered in [Maire & Youk, 2015a], but its results were limited to an extension of the phenotype diagram calculation to systems with a finite Hill coefficient. Here, we will examine more thoroughly the dynamics and steady states of the extended model with sigmoidal response. We start by explaining our main qualitative findings in Section 3.2.2. In Section 3.2.3, we derive exact solutions of a one-dimensional model describing both the steady states when the cells interact strongly and the lattice homogenizes, as well as when they interact weakly and can be treated as autonomous. Section 3.2.4 discusses the transition from this fully homogeneous phase to an autonomous phase that occurs by modulating the density of cells. Finally, in Section 3.2.5 we study aspects of the bistability of our system by deriving a mean-field description.

### 3.2.1. MODELING CELLS WITH CONTINUOUS RESPONSE CURVES

Our starting point is the model introduced in the previous section, which we extend to consider a continuous response function (Hill function) as we will describe next. In the continuous picture, the secretion rate of a cell is a continuous variable which we assume to lie between a lower bound  $C_{OFF}$  and an upper bound  $C_{ON}$ . Hence, the state of the cell is specified by a continuous variable  $X \in [0, 1]$  that scales linearly with the secretion rate. More precisely, the secretion rate and the state of the cell are related through the relation

$$C(X) = C_{OFF} + (C_{ON} - C_{OFF})X. \quad (3.9)$$

The sensed signaling molecule concentration of a cell  $k$  is the sum over the contributions from each of the cells in the system, including cell  $k$  itself. The contribution from the cell

itself is equal to its secretion rate  $C(X_k)$ , whereas the contribution of a cell  $l$  at distance  $r_{kl}$  equals  $f(r_{kl})C(X_l)$ , where  $f(r_{kl}) = \frac{e^{R-r_{kl}}}{r_{kl}} \sinh(R)$ . Altogether, this means that we can write the sensed concentration of cell  $k$  as

$$Y_k = C(X_k) + \sum_{l \neq k} f(r_{kl})C(X_l). \quad (3.10)$$

In response to the sensed concentration, the cell updates its secretion rate according to the aforementioned Hill function,

$$C(X_k(t+1)) = C_{OFF} + (C_{ON} - C_{OFF}) \frac{Y_k(t)^n}{K^n + Y_k(t)^n}, \quad (3.11)$$

Note that this update function is a sigmoid that varies between  $C_{OFF}$  (when the cell senses nothing,  $Y_k = 0$ ) and  $C_{ON}$  (when the sensed concentration is infinite). Together with the definition of the cell state, this implies that cell states evolve as

$$X_k(t+1) = \frac{Y_k(t)^n}{K^n + Y_k(t)^n} \quad (i = 1, 2, \dots, N). \quad (3.12)$$

In the limit when  $n \rightarrow \infty$ , the Hill function reduces to a Heaviside step-function and the original system with a binary response is recovered:

$$X_k(t+1) = \theta(Y_k(t) - K). \quad (3.13)$$

In the case when the feedback of the signaling molecule is negative, i.e. when the sensed molecule represses its own production, the dynamics is described by a negative Hill function,

$$X_k(t+1) = \frac{K^n}{K^n + Y_k(t)^n} \quad (i = 1, 2, \dots, N). \quad (3.14)$$

Note that this corresponds to a decreasing sigmoid function for the secretion rate as a function of the sensed concentration, with maximum  $C_{ON}$  (when  $Y_k = 0$ ) and minimum  $C_{OFF}$  (when  $Y_k \rightarrow \infty$ ). In the following, we only consider the positive feedback case, but the case of negative feedback produces interesting oscillatory phenomena and allow for further studies of the model.

We performed numerical simulations of the continuous model with random initial conditions, where we randomly chose the initial state of each cell to be uniformly distributed in the interval  $[0, 1]$ . Other schemes for initiating the system will be discussed in later sections. The equilibrium condition  $X_k(t+1) = X_k(t)$  for all cells  $k$  is typically only reached asymptotically as  $t \rightarrow \infty$ . Therefore, we terminated the simulation whenever

the changes in cell states become smaller than some threshold  $\varepsilon$ , i.e. when

$$\forall k : |X_k(t+1) - X_k(t)| < \varepsilon. \quad (3.15)$$

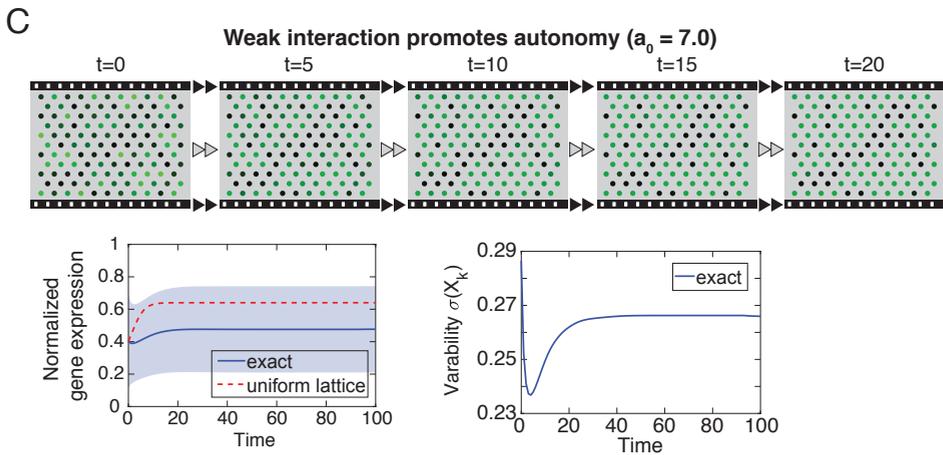
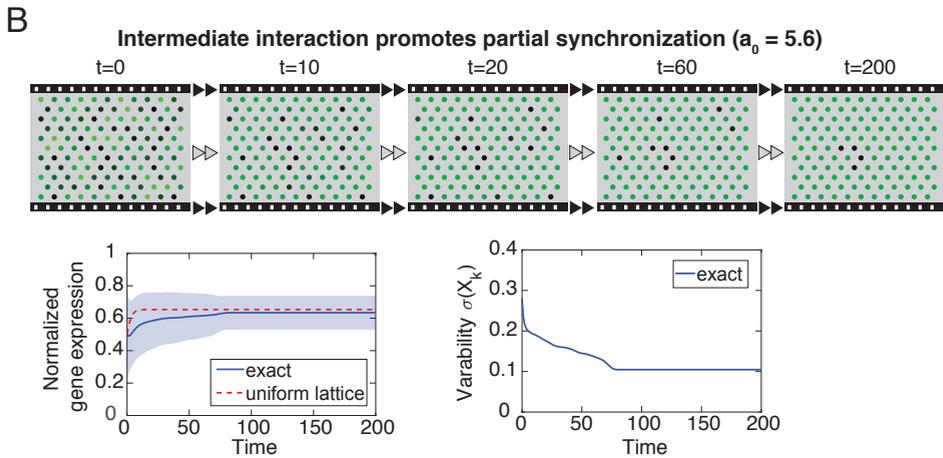
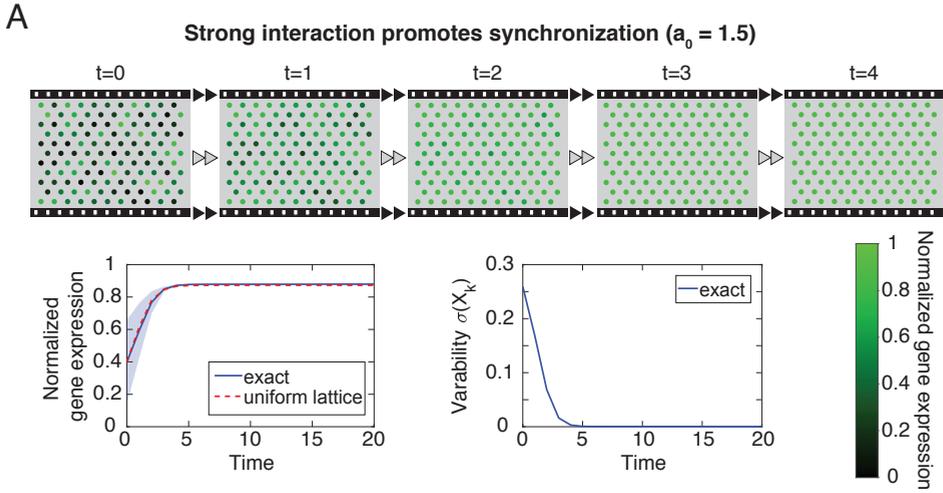
What is the interpretation of the value of this  $\varepsilon$ ? If the difference between two single-cell states is  $\varepsilon$ , then the difference in expression value between the states is  $(C_{ON} - C_{OFF})\varepsilon$ . Hence  $\varepsilon$  gives the minimum change in expression level we allow, in units of  $(C_{ON} - C_{OFF})$ . In practice, we typically set  $\varepsilon = 10^{-5}$ . In the following, we also set  $C_{OFF} = 1$ , such that all concentrations are measured in units of the 'OFF' secretion rate.

### 3.2.2. COLLECTIVENESS VERSUS AUTONOMY

Through performing parameter sweeps on our model, we observed that our system has two distinct phases in which the dynamics simplifies, such that the system becomes easy to characterize. In all of the following, we will take a Hill coefficient of  $n = 2$ , because in this case the system is bistable both at the single-cell level and population level for a range of parameters, as will be shown in the next section. However, our qualitative findings hold for any value of the Hill coefficient which is sufficiently low, such that the response function is significantly different from a step function. As such, we can identify "ON" and "OFF" states in analogy to the infinite Hill coefficient model and more directly compare their features.

The phase space location of the two phases are mostly determined by the variable  $a_0$ , which represents the distance between neighboring cells in units of the signaling length. Note that  $a_0$  is inversely related to the density of the cells. At low  $a_0$  (high density), communication between cells is strong, causing the cells to synchronize their gene expression levels. We shall refer to this regime as the *collective phase* or the *strong interaction phase*. At high  $a_0$  (low density), communication between cells is weak and cells individually tune their gene expression over time. We shall refer to this regime as the *autonomous phase* or *weak interaction phase*. As  $a_0$  is gradually modulated between these extremes, we find a gradual transition from one to the other type of behavior with interesting features in between.

At very low  $a_0$  (strong interaction), we observed that the lattice of cells always evolved towards a uniform state where all cells acquired identical states, as shown in the example of Figure 3.2A. Starting from a disordered initial configuration where each cell has a random state, we observed that the average gene expression level approached a constant value, while the variation between cell states uniformly decays to zero. Hence the gene expression levels of different cells on the lattice become identical to each other. Mathe-



matically, this homogenization is a consequence of the fact that cells are able to respond to arbitrarily small changes in the model. In the binary model, the cells only respond to small differences in sensed concentration, if the concentration is near the sensing threshold  $K$ . However, with a continuous response function, arbitrarily small changes in the sensed concentration affect the cell's state (Eq. 3.12). This allows the cells to finely tune their state to match that of their surroundings. Note that in real biological systems, sensitivity to sensed concentrations cannot be arbitrarily small, as there are fundamental limits to how accurately a cell can sense [Berg & Purcell, 1977]. Such effects could be dealt with by introducing a stochasticity at various parts of the model, but we shall not discuss this in detail.

As  $a_0$  is increased, we reach a transition regime in which there is some collective dynamics, but the system may not become fully homogeneous (Fig. 3.2B). In the example, we observe that most of the number of cells turn "ON" over time, but at the end of the simulation a cluster of four "OFF" cells remain. The overall gene expression level increases slightly, while the variation between cells goes to a constant non-zero value. This is the only phase in the model in which we observed non-trivial pattern formation, in which the system becomes more spatially correlated over time. This is the result of two opposing effects balancing each other out: the tendency for lattices to homogenize if the cells would couple more strongly and the tendency for cells to be unaffected by neighbors if they would couple more weakly. In this intermediate regime, only a subset of the cells switch their state and this depends on the states of their neighbors. The result is the formation of island patterns of gene expression similar to what has been observed in the binary model.

As  $a_0$  is further increased to high values (weak interaction), we also observed simulations where the cells largely evolved independently, similar to what would happen if the cells did not communicate with each other. In this case, each cell's state reaches to a steady state that depends mostly on the cell's initial states, rather than on the states of other cells in the system. As such, we observe that the mean gene expression level also approaches a constant value, while the standard deviation of the cell states approaches

---

**Figure 3.2 (preceding page): From collectiveness to autonomy: example simulations.** We show simulations at different values of the inter-cell spacing  $a_0$ , which is a measure for the strength of the interaction between the cells. The filmstrips show snapshots from the simulation. Greener colors correspond to higher levels of gene expression (see colorbar). Plots below the filmstrips show the average and spread of the cell states (gene expression level) of the system over time. The variability of the lattice gene expression is defined as the standard deviation of the cell states. The shaded area indicates this variability, i.e. they mark values within one standard deviation of the mean. Results are for exact simulations (blue) and for a numerical solution of the 1D uniform lattice system initiated with the same average gene expression value (red dotted line). (A) Fast synchronization in the strong interaction regime ( $a_0 = 1.5$ ). (B) Slower and incomplete synchronization in the intermediate interaction regime ( $a_0 = 5.6$ ). (C) Autonomy in the weak interaction regime ( $a_0 = 7.0$ ).

a relatively high positive value (Fig. 3.2C). Qualitatively, the dynamics in the weakly interacting regime is similar to what would happen if the cells did not communicate at all.

In the collective phase, the dynamics of the system reduces to that of a one-dimensional system. Since the cells are placed on a regular lattice, each cell of a uniform lattice senses the same concentration, and therefore a uniform lattice will remain uniform in the absence of noise. The state of the uniform lattice is fully described by a single variable  $X_m \equiv X_1 = \dots = X_N$ , and its time evolution is given by

$$\begin{aligned} X_m(t+1) &= \frac{Y_m(t)^n}{K^n + Y_m(t)^n} \equiv f_m(X_m(t)) \\ Y_m(t) &\equiv (1 + f_N) [X_m(t)(C_{ON} - 1) + 1]. \end{aligned} \quad (3.16)$$

Note that this is equivalent to the time evolution map of a single cell, up to a multiplicative factor of  $(1 + f_N)$  for  $Y_m(t)$ . This means that the dynamics of a uniform lattice is equivalent to that of a single lattice, up to a rescaling of the parameters. The required rescaling is obtained through a single rescaling of the threshold  $K$ :

$$K' = \frac{K}{(1 + f_N)^{1/n}}. \quad (3.17)$$

Hence, the solutions of Eq. 3.16 for a uniform lattice with parameter  $K$  are equivalent to those of a single cell with a parameter  $K'$ . Note that if the interaction is weak ( $f_N \ll 1$ ), then  $K' \approx K$ . Therefore, the steady states of Eq. 3.16 also well-approximate the single-cell steady states in the weakly interacting regime. In this case, rather than having one equation describing the entire lattice, our system decouples into  $N$  identical equations describing each of the individual cells.

We first note that the dynamics of the uniform lattice model approximates the dynamics of the average gene expression level of our full system in the strong interaction phase (Fig. 3.2). More specifically, in this case the dynamics is in excellent agreement with the simulated average gene expression level (Fig. 3.2). In particular, the both equations evolve to the same steady state value. At intermediate interaction, the transients of the two trajectories deviate, but the steady state values still closely approach each other, since the final configuration is nearly uniform (Fig. 3.2B). Finally, in the weak interaction regime, the agreement between the trajectories is poor (Fig. 3.2C). This is because the uniform lattice solution represents the dynamics of a single cell (recall that  $K' \approx K$  for weak interaction) and evolves toward one of the single cell steady states, but since

there are two possible steady states for every cell, the population average is in between these single cell steady states.

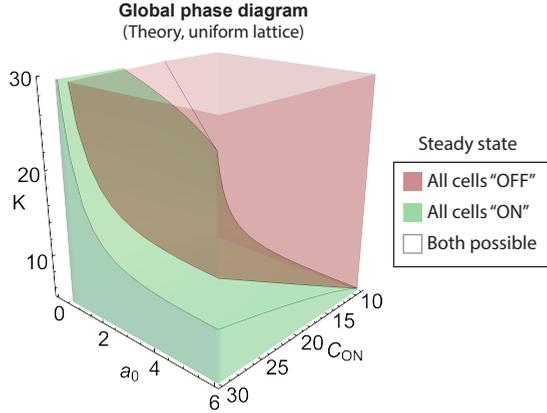
Next, we studied the solutions of the one-dimensional model by determining the number, location and stability of the fixed points of this system. For the Michaelis-Menten case of  $n = 1$ , it is straightforward to show that there can only be one stable fixed point whose value varies from low to high as parameters of the system are changed (see Section 3.2.7). For moderate ultrasensitivity at  $n = 2$ , this revealed three qualitatively different ‘phases’, whose boundaries depend on the values of  $a_0$ ,  $K$  and  $C_{ON}$  (as well as  $f_N$ , which is a function of both  $a_0$  and  $N$ ) (Fig. 3.3). Two of the phases correspond to monostable steady states where the uniform lattice had only one solution. For values of  $K \ll C_{ON}$ , there is a single solution at a high value of  $X_m$  — which we identify as the “ON” state of the system. In the other extreme, when  $K \gg C_{ON}$ , the system always evolved toward a fixed point at low  $X_m$ , which we identify as the “OFF” state of the system. For intermediate values between these extremes, there is a bistable phase where both the “ON” and the “OFF” states are stable. The boundaries of the bistable region depend on the value of  $a_0$ . For lower values of  $a_0$  (stronger interaction), the boundaries will have relatively higher values of  $K$  at any value of  $C_{ON}$ , since a strong interaction requires a higher threshold to reach the same effect. These boundaries can be computed analytically for  $n = 2$ , as we show in Section 3.2.7. Furthermore, we checked that the solutions of the one-dimensional system have the same stability as uniform lattice solutions in the  $N$ -dimensional system (Appendix 3.2.7). Hence, small perturbations to a uniform lattice steady state will always take the system back to a uniform lattice configuration.

We then compared the findings of the uniform lattice model to batch simulations where we ran many simulations for a fixed set of parameters with different initial states. We did this at fixed values of  $a_0$  in the strongly interacting regime and the weakly interacting regime. In the strong interaction phase ( $a_0 = 1.5$ ), the bistable region lies in a narrow region slightly above the line  $K = C_{ON}$  and starts at intermediate values of  $K$  and  $C_{ON}$ , indicating that for very low values of either  $K$  or  $C_{ON}$  there is no bistability. The steady state solutions of the system show an interesting bifurcation diagram (Fig. 3.4B), which shows how the steady states change as the parameters of the system are altered. Let us fix  $C_{ON}$  and discuss what happens as  $K$  is increased from low to high values. At a low  $C_{ON}$ , the system is always monostable, and we find that increasing  $K$  decreases the value of the steady state  $X_{SS}$  at the single fixed point with a rather sharp transition between the low and high values. At higher values of  $C_{ON}$ , we find a bistable region flanked by two bifurcations as  $K$  is increased across the range of examined values. At high values of  $K$ , there is a single steady states with  $X_{SS} \approx 0$ . At low values of  $K$ , there is a single steady

state with  $X_{SS} \approx 1$ . As  $K$  is modulated between these extremes, the two solutions coexist in an intermediate region through the appearance of an additional unstable fixed point. As  $K$  is increased, first a saddle-node bifurcation marks the appearance of a low fixed point and intermediate unstable state. As  $K$  is further increase, another saddle-node bifurcation occurs as the unstable state joins the high fixed point and both are annihilated, so that only the low fixed point persists. A consequence of this bifurcation diagram is the appearance of hysteresis, marked by a history-dependence of the steady state as the parameters are modulated across the bistable region. Here, it implies that if  $K$  is gradually increased from very high value to some intermediate value, the system would remain stable at a relatively high fixed point, until it makes a sudden jump to the low fixed point. If  $K$  is gradually decreased from a very low value up to the same intermediate value, the system would remain at a relatively low fixed point, before making a discontinuous jump to the high fixed point as  $K$  is increased further. From our batch simulations, we found that the system indeed always becomes uniform across the examined parameter range (data not shown). The fixed points also correspond nicely to those computed from the uniform lattice model, with a region of bistability as predicted (Fig. 3.4D). Note that the intermediate fixed points are missing as they correspond to unstable steady states and are practically never observed in simulations.

### 3.2.3. EXACT SOLUTIONS OF A ONE-DIMENSIONAL MODEL

In the weakly interacting regime, the phase diagram computed in Fig. 3.3 applies not to the entire lattice, but to individual cells of the system. In the monostable regions, individual cells are monostable, and therefore the entire lattice will also have a single fixed point where all the cells are identical. However, this occurs not due to synchronization between cells, but because each individual cell evolves towards the fixed point. In the bistable phase, since each of the cells is bistable, there are  $2^N$  possible steady states of the system. As a function of  $K$  and  $C_{ON}$ , this bistable phase corresponds to a similarly shaped narrow region as the bistable region for strong interaction (green region in Fig. 3.4E). The simulations confirm that outside of this region, the lattice always evolves to the same final state (Fig. 3.4F). In contrast, inside the bistable region, we uncovered a range of possible final states (with different values of average gene expression level  $\langle X_k \rangle$ ) as shown in Fig. 3.4F). Altogether, these results confirm that we can use Eq. 3.16 to study the steady states of the system in the collective phase and in the autonomous phase.

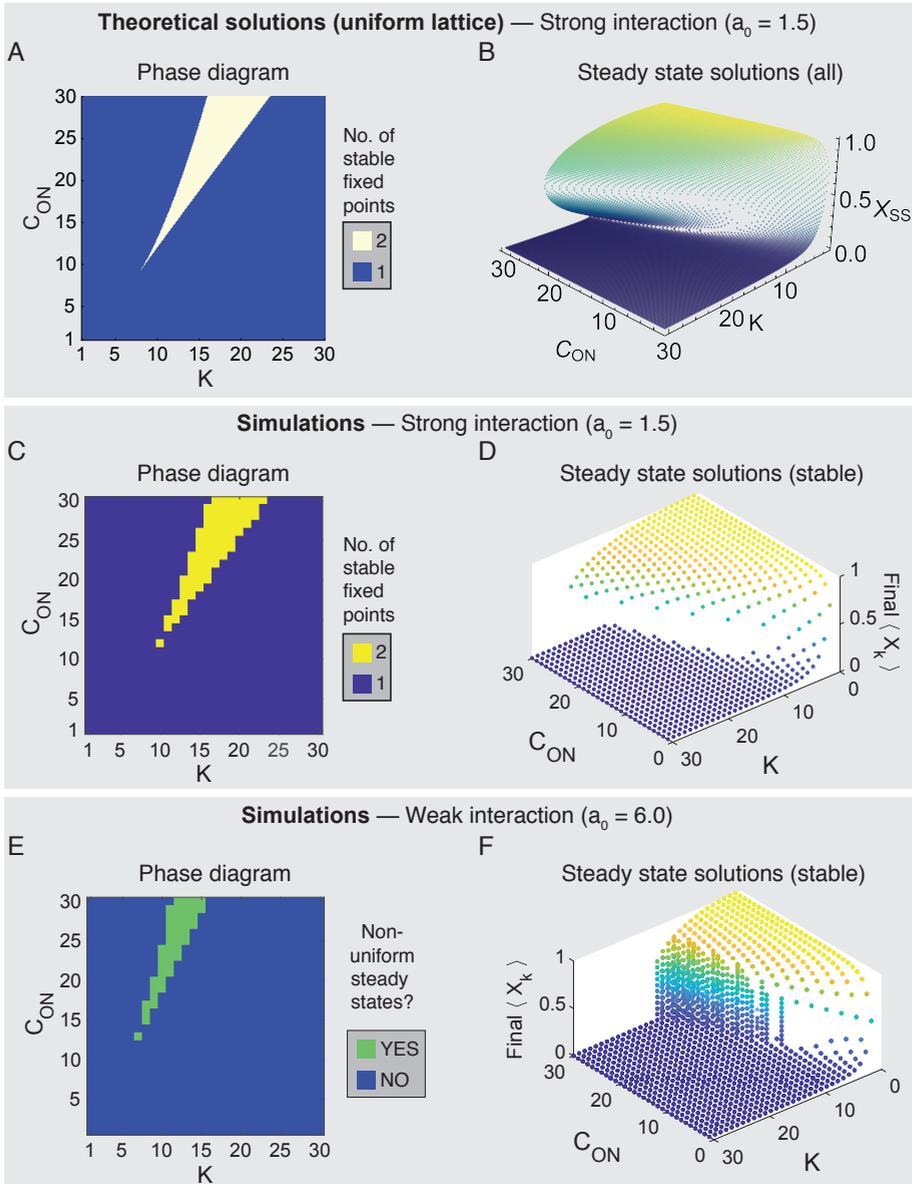


**Figure 3.3:** Global phase diagram shows the possible steady states of a uniform lattice system, determined by solving the one-dimensional model given by Equation 3.16. In the red region, there is a single steady state corresponding to a low expression level for all cells. In the green region, there is also a unique steady state corresponding to a high expression level. In the region between the red and green regions, the system is bistable. Axes variables are the distance between neighboring cells  $a_0$ , the ON-cell secretion rate  $C_{ON}$  and the sensing ON-threshold  $K$ .

### 3.2.4. TRANSITION FROM COLLECTIVE TO AUTONOMOUS DYNAMICS

We next studied how the system changes from the collective synchronization regime to the autonomous regime as we gradually modulate the parameters of the system (Figs. 3.5, 3.6). We did this by changing  $a_0$  at fixed value of  $K, C_{ON}$ , such that in the collective phase it is monostable (the system homogenizes and always reaches the same steady state) whereas in the autonomous phase it is bistable (each individual cell has two stable fixed points). By tuning  $a_0$  between these extremes, we found a transition region in which the system shows more complicated dynamics that cannot be predicted from the one-dimensional model only. In particular, we observed that the cells were able to partially synchronize, forming islands of cells with similar gene expression, but that these islands often could not spread across the entire system. This type of dynamics is similar to what happens in the active phases (activation/deactivation) of the binary model. As such, it seems to be the only regime in the continuous response model for which non-trivial pattern formation is possible.

Our next goal was to identify variables — analogous to order parameters in statistical physics — which show how features of the system change as the system transitions from the collective phase to the autonomous phase. Because the trajectories in the transition phase show considerable variability depending on their initial configuration, the ‘order parameters’ we examined correspond to statistical quantities based on results of many

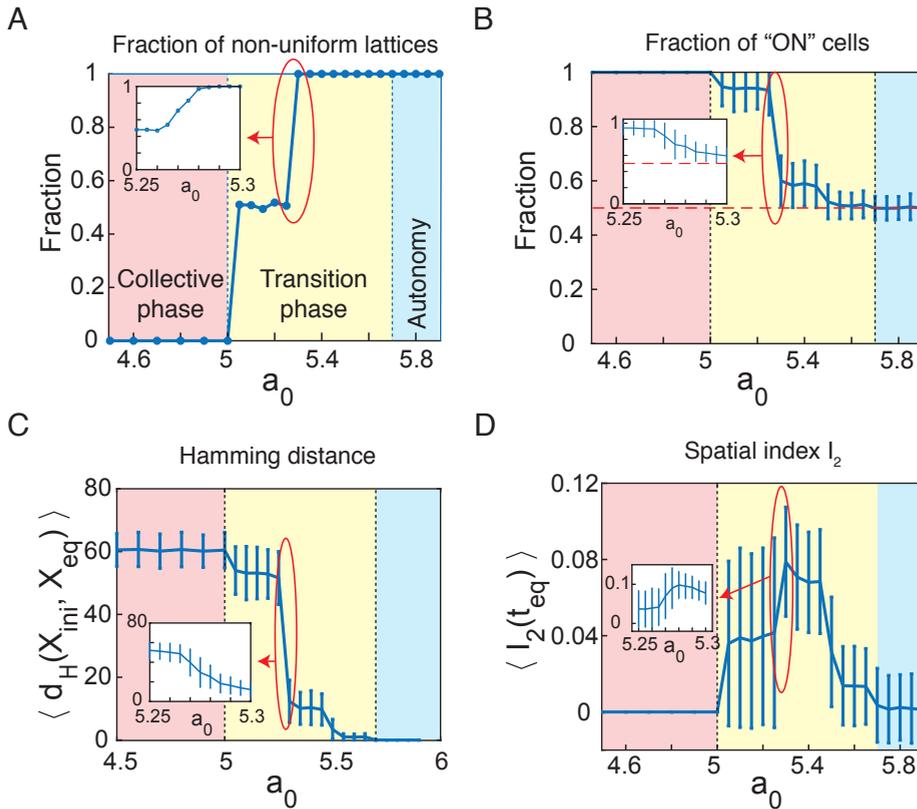


**Figure 3.4:** Steady state phase diagrams obtained from theory and simulations. (Caption continued on next page)

**Figure 3.4 (previous page):** (A-B) Theoretical results in the strong interaction regime ( $a_0 = 1.5$ ). (A) Number of fixed points as a function of  $K$  and  $C_{ON}$ . (B) Steady state solutions across the studied range of  $K$  and  $C_{ON}$ . For some values of  $K, C_{ON}$ , there is a unique solution, whereas for other there are three solutions. In the latter case, the middle solution is always unstable. (C-F) Simulation results in the strong and weak interaction regimes. For each value of  $(K, C_{ON})$ , we chose initial mean gene expression level  $\langle X_k \rangle(t=0) = 0, 0.1, \dots, 1.0$  and performed 10 simulation for each  $\langle X_k \rangle(t=0)$ . (C-D) Strong interaction regime ( $a_0 = 1.5$ ). For each of the simulations, the cells synchronized their gene expression levels such that the final state is always a homogeneous lattice. (C) Number of fixed points for every fixed value of  $K, C_{ON}$ . (D) Location of the stable fixed points. (E-F) Weak interaction regime ( $a_0 = 6.0$ ). We observed the existence of autonomous steady states, which are characterized by non-uniform steady states where individual cells obtain different final states mostly due their individual dynamics rather than due to neighbor-induced transitions. (E) Phase diagram shows for which values of  $K, C_{ON}$  non-uniform steady states are possible. (F) Location of the stable fixed points.

simulations. In the following, we shall discuss four such quantities, all of which seemed to show a continuous trend as  $a_0$  was varied across the transition regime. For each of the quantities, we discuss their ability to capture the main features of the ‘phase transition’. Three of the quantities are also dynamical quantities that can be studied as a function of time, and we will discuss their dynamics alongside their final values used to characterize the transition.

**Fraction of non-uniform lattices** The first measure we considered was the fraction of non-uniform lattices. For each value of  $a_0$ , we ran a large number of simulations and counted the fraction of simulations with final states where the cells did not have (nearly) identical gene expression levels. More precisely, we defined a lattice as non-uniform whenever the standard deviation between the cells was smaller than some threshold  $\varepsilon$ . In the collective phase, all lattices are uniform and this fraction equals to zero. In the autonomous phase, if we start with random initial configurations, we expect none of the lattices to become uniform. As  $a_0$  is increased, we observe a sharp increase from 0 in the collective phase to a value close to 1/2 in the transition region (Fig. 3.5A). Upon further increasing  $a_0$ , the metric sharply increases to one. Upon closer inspection, it appears that the metric varies smoothly between these extremes (see inset of Fig. 3.5A). However, the plateauing value of 1 is reached before the onset of the autonomous phase identified from other metrics. This is because the system does not need to be fully autonomous (such that each cell’s final state is determined only by its own initial state) in order to produce non-uniform final states. The early plateauing of the metric in the transition phase only shows that within this regime, the system never fully synchronizes anymore. In short, the fraction of non-uniform final state lattices is an intuitive metric that tells us whether the system reaches a homogeneous final state or not. However, its main drawback are that it does not directly quantify the degree of autonomy of the system is and that the findings depend on the initial states of the simulations chosen.



**Figure 3.5: Transition between the collective phase and the autonomous phase.** (A-D) Various measures characterize the transition to autonomy as the inter-cell distance  $a_0$  is gradually increased. At low  $a_0$ , we have a collective phase where the system always synchronizes to become uniform (red). At high  $a_0$ , each cell largely follows its intrinsic dynamics regardless of what the rest of the system does (blue). We focused on the transition region in between these extremes (yellow). For each value of  $a_0$ , we performed 500 simulations with cells that have state  $X_k = 0$  or  $X_k = 1$  with probability  $1/2$  each. Insets show a zoomed in region ( $5.25 \leq a_0 \leq 5.3$ ) where the measures change abruptly. (A) The probability to generate a non-uniform steady state becomes non-zero outside the collective phase. We define non-uniform steady states as those for which the standard deviation of the cell states in steady states is smaller than some threshold  $\epsilon = 0.01$ . (B) The fraction of cells which have a final 'ON' state in the digital representation shows a sudden drop at the autonomy transition. (C) The Hamming distance between the digitized cell states of the initial and final states is a measure for how much the system has evolved, and precisely marks the onset of the autonomy phase. (D) The modified spatial index  $I_2$  (Eq. 3.20) acquires non-zero values on average in the transition region, where pattern formation is possible.

**Fraction of "ON" cells** The next metric we considered is the fraction of "ON" cells in a binary description of the system's states. Since we are working in the bistable regime of the  $n = 2$  model, we could effectively treat the cells as binary in a static description of their state, so we can apply the earlier developed tools and concepts for binary cells to describe these cells. In particular, one useful quantity to consider is the fraction of

"ON" cells in the steady state of the system. Below, we first formally construct a binary description and then examine how this fraction of "ON" cells changes as  $a_0$  was tuned across the transition region.

We defined a cell to be "ON" if its expression is higher than that of the unstable fixed point (for a uniform lattice), and "OFF" if it is lower. Let  $X_u^*$  be the unstable fixed point of the uniform lattice model (Eq. 3.16). Formally, given a state  $X \in [0, 1]^N$ , we then define a corresponding state  $\tilde{X} \in \{0, 1\}^N$  by converting each cell into a binary variable. Define a function  $c = (c_1, c_2, \dots, c_N) : [0, 1]^N \rightarrow \{0, 1\}^N$ , which on each component acts as

$$c_i(X_i) = \theta(X_i - X_u^*). \quad (3.18)$$

In reality the unstable fixed point and basins of attraction for an interacting system of cells may not precisely correspond to that of the uniform lattice system. Nevertheless, as a first approximation, for cells whose state is not close to  $X_u^*$  this approximation should be sufficient to tell whether they remain in their original state or transition to the other state. Furthermore, this approach can be extended straightforwardly to systems with more than two single-cell fixed points. This occurs in systems with higher Hill coefficient. If the single-cell system has  $m$  fixed points, separated by  $m - 1$  unstable fixed points, then we identify  $m$  states  $\{0, 1, \dots, m\}$ , which we order according to their  $X$  value. We would then map each cell onto one of these  $m$  states depending on which basin of attraction it falls in.

The fraction of "ON" cells of the steady state always takes a value of either 0 (all "OFF") or 1 (all "ON") in the collective phase. In the autonomous phase, the fraction equals that of the initial configuration of the system. Since we independently choose each cell's gene expression with  $P[X_k = 0] = P[X_k = 1] = 1/2$ , our final state fractions are also binomially distributed with parameters  $(N, p = 1/2)$  and in particular with mean  $1/2$ . Furthermore, this metric seems to vary smoothly between these extremes as  $a_0$  is increased (Fig. 3.5B). Just like the previous metric, this metric precisely determines the onset of the transition phase from the collective phase, but does not precisely determine the onset of the autonomous phase. The onset of the transition phase can be defined as the first value of  $a_0$  where we observe a trajectory where not all cells are "ON" or "OFF" anymore. As such, it's identical to the fraction of non-uniform lattices for determining the onset of this transition phase. At the other end of the transition phase, the metric gradually shifts to a value of  $1/2$ . However, because also in the autonomous phase the fraction has a spread around its average value of  $1/2$ , it is hard to precisely pinpoint where this region starts, unless one exhaustively simulates all states of the system.

We also examined the dynamics of the fraction of "ON" cells across this transition range (Fig. 3.6A). Our first observation is that with the set of parameters chosen, the cells only

seemed to be able to activate, i.e. switch from “OFF” to “ON”, but not vice versa. Intuitively, this is because of the asymmetry of the “OFF” and “ON” states: the “ON” state has a far larger basin of attraction and hence it is easier to collectively activate than to deactivate. We further note that with similar initial conditions, the individual trajectories show considerable heterogeneity, just as in the binary model. Let us now examine what happens as  $a_0$  is increased in more detail. At  $a_0 = 5.0$  (collective phase), all cells activate over time, such that the fraction of ON cells always reaches 1 eventually. As we reach the transition regime, at  $a_0 = 5.4$ , the system no longer fully activates, but reach an equilibrium after a fraction of the cells has been activated. Upon further increasing  $a_0$  to 5.6, we see very few activation events. What is surprising here is that the activation of a single or a few cells now takes a substantial amount of time (around 60-80 time steps), indicating at slow dynamics whereby the cells states change by small amounts at each time step. Finally, in the autonomous regime at  $a_0 = 6.0$ , none of the trajectories shows any activation and the fraction of “ON” cells remains constant over time.

**Hamming distance** The next measure we considered describes how much the initial and final configurations in our system differ from each other. This is directly related to the degree of autonomy in the system. In the autonomous state, cells never transition between different basins of attraction, so in the binary picture, there is no difference between the initial and final states. In the collective phase, depending on the initial condition, possibly many cells switch state to be in tune with their neighbors, and the initial and final states differ considerably. To quantify this difference, we considered a distance function  $d_H : [0, 1]^N \times [0, 1]^N \rightarrow \mathbb{R}$  between two states  $X$  and  $Y$

$$d_H(X, Y) = \sum_{i=1}^N |c(X_i) - c(Y_i)| \quad (3.19)$$

Hence, we consider the difference between the digitized states of two different configurations. This metric is known as the Hamming distance and has been proposed to be a suitable metric for measuring distances between different states in a cellular automaton [Wolfram, 1983]. We applied this metric to measure the difference between the initial state  $X_{ini}$  and final state  $X_{eq}$  of the system. In the collective phase, on average we have  $\langle d_H(X_{ini}, X_{eq}) \rangle = N/2$  with our choice of initial conditions, since roughly half of the cells have each of the two binary states initially, but as the system homogenizes, all cells within one of these groups change to the other state. In the autonomous phase,  $d_H(X_{ini}, X_{eq}) = 0$  by definition, since none of the cells should change their basin of attraction and jump to the other fixed point. Hence, the Hamming distance provides an exact definition of the autonomous phase. We see that its value also shows a monotonic, continuous trend from  $N/2$  to 0 upon increasing  $a_0$  (Fig. 3.5C). Close to the autonomy

phase border, the Hamming distance takes very small but non-zero values. This is when only a small fraction of cells within each simulation changes their state. Nevertheless, this indicates that the system is not fully autonomous yet, and therefore the Hamming distance can precisely indicate the onset of autonomy in the system.

The dynamics of the Hamming distance contains comparable information as the fraction of “ON” cells (Fig. 3.6B), with a few notable differences. The Hamming distance does not distinguish between activation and deactivation, so each change in value could be both due to an “ON” cell turning “OFF” as well as an “OFF” cell turning “ON”. Furthermore, in the collective phase, the maximum value the Hamming distance can reach depends on the initial fraction of “ON” cells, which is why the trajectories in the  $a_0$  plot do not plateau at the same levels. Without additional information about the initial state, we cannot therefore not tell whether the trajectories has "maximally diverged" from its initial state.

**Modified spatial index** Since the transition region shows examples of trajectories that become spatially organized, we sought to define a metric to quantify the degree of spatial organization of the patterns generated. The previously defined spatial index (Eq. 4.1 in Section 4.2.2) does not generalize well to our continuous system. This is because the spatial index is undefined for a uniform lattice. However, this can be fixed by removing the variance from the denominator, so that the new quantity becomes

$$I_2 = \frac{\Theta}{f_N} - (2\langle X_k \rangle - 1)^2, \quad (3.20)$$

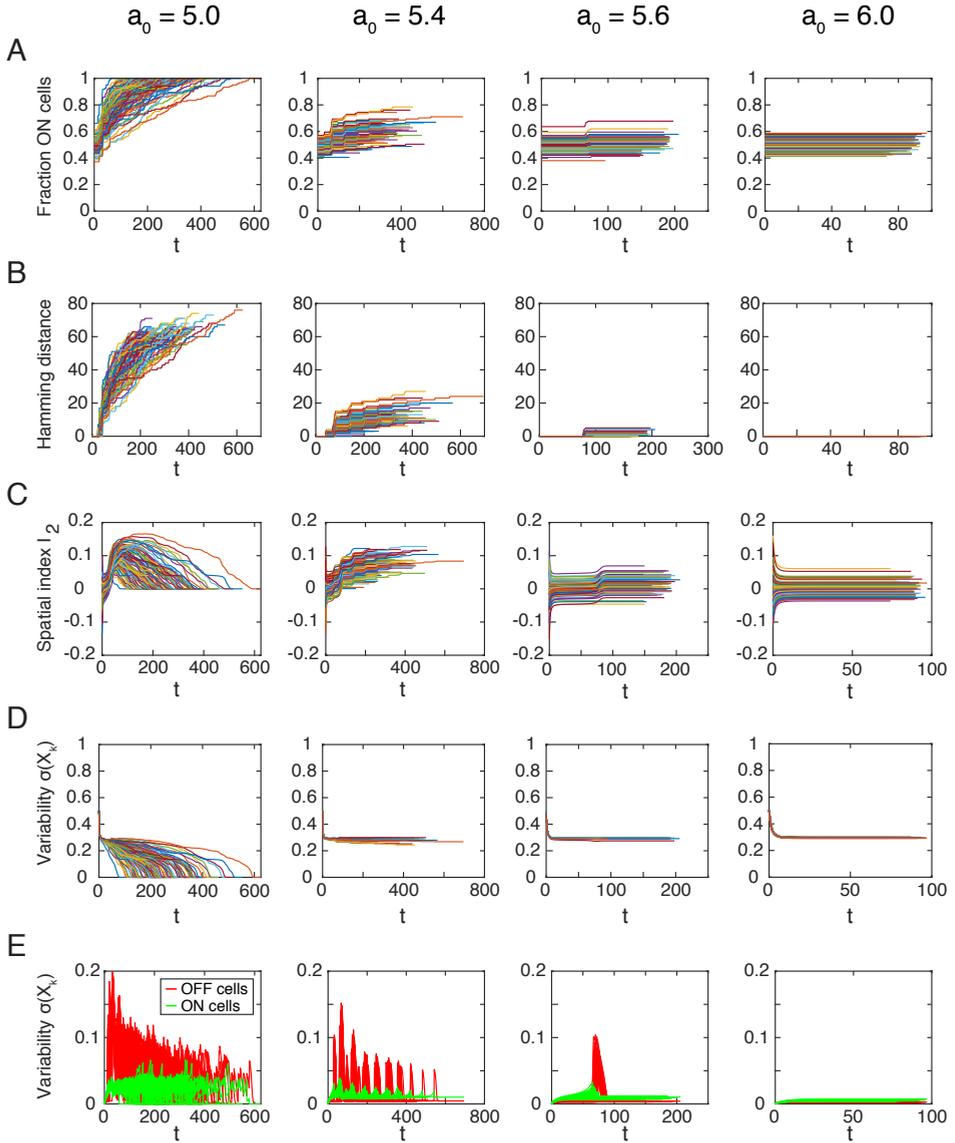
where  $\Theta = \frac{1}{N} \sum_{k,l} f(r_{kl}) \tilde{X}_k \tilde{X}_l$  (with  $\tilde{X} \in [-1, 1]$  defined as  $\tilde{X} = 2X - 1$ ). The downside of leaving out the variance is that it becomes hard to interpret the numerical value of  $I_2$ . However, our main goal here is to distinguish homogeneous and disordered lattices — both of which have  $I_2 \approx 0$  here, from spatially organized patterns, which have  $|I_2| > 0$ . To see this, note that for a homogeneous lattice  $\Theta = f_N (2\langle X_k \rangle - 1)^2$ , and therefore  $I_2 = 0$ . In the autonomous phase, we generate random initial configurations that tend to be spatially disorganized. Since our metric differs from the previous metric only by a constant, these configurations should also have  $I_2 \approx 0$ . However, as seen from Fig. 3.5D, there is considerable variability in the autonomous phase. Therefore, also the modified spatial index cannot precisely indicate the onset of the autonomous phase, while it can precisely indicate the end of the collective phase. In the transition region, the pattern formation discussed earlier leads to higher values of the spatial index than in the other phases. Dynamically, the modified spatial index shows very different time traces for the different phases of the system (Fig. 3.6C). In the collective phase ( $a_0 = 5.0$ ),  $I_2$  first increases

rapidly as “OFF” cells close to “ON” cells increase their gene expression level, since this drives the formation of islands of cells with similar gene expression. However, as more and more cells turn “ON”, the lattice becomes more homogeneous over time and the  $I_2$  gradually decays to zero. In contrast, in the transition region ( $a_0 = 5.4$  and  $a_0 = 5.6$ ) only a few cells activate, which tends to increase the spatial index over time without ever reaching that critical mass of “ON” cells at which  $I_2$  starts decreasing. Finally, in the autonomous phase ( $a_0 = 6.0$ ) the spatial index rapidly reaches a constant value as individual cells reach their steady state values, and remains constant and close to zero over time.

In summary, we have defined different metrics for distinguishing between the collective phase and the autonomous phase in the system. Three of these metrics — the fraction of non-uniform final states, the mean fraction of “ON” cells in the final state and a modified spatial index — precisely determine the boundary between the collective phase and the transition region. The remaining metric — the Hamming distance between initial and final state — precisely determines the boundary between the transition region and the autonomous phase. Apart from these metrics, there are more metrics that can be used either as substitute for some of the metrics we discussed or to provide additional information. One such metric is the variability in gene expression of the system, quantified by the standard deviation of the cell states  $\sigma(X_k) = \sqrt{\langle X_k^2 \rangle - \langle X_k \rangle^2}$ . This variability goes to zero if and only if the system becomes homogeneous, and is therefore a direct measure for the degree of homogeneity in the system (Fig. 3.6D). A homogeneous system requires that the cells in the system are either all “ON” or all “OFF”, but the reverse does not need to be true: if all cells had the same binary states, there could still be considerable variation in their gene expression levels. This is because the binary states are defined in terms of the basins of attraction, which have a finite extent. To study the variability among the groups of “ON” and “OFF” cells in the autonomous system, we separately examined the standard deviation of the cells states of the “ON” cells and that of the “OFF” cells. This showed that after a transient phase, there was very limited variability among the cells within each of both groups (Fig. 3.6E). Why do the cells maintain their binary character in this transition phase? Why do we hardly find any steady states with cells whose state deviates significantly from one of the single-cell fixed points? To address these questions, we resorted to a different theoretical description.

### 3.2.5. PERSISTENCE OF BISTABILITY

To explain the binary character of the steady states (Fig. 3.7A), we resorted to a reduced description of the system which facilitates analysis of the steady states while preserving



**Figure 3.6: Dynamics of various metrics describing trajectories in and close to the transition region between the collective and the autonomous phases.** We plot results from 100 simulations for values of  $a_0 = 5.0$  (collective phase),  $a_0 = 5.4$ ,  $a_0 = 5.6$  (transition phase), and  $a_0 = 6.0$  (collective phase). The simulations were initiated with random states where each cell had probability 1/2 to have state 0 and probability 1/2 to have state 1. (A) The fraction of cells identified as ‘ON’, i.e. which are in the basin of attraction of the higher steady state (see Eq. 3.18). (B) Hamming distance between the binary states of at time  $t$  and the initial state at time 0. (C) Spatial index  $I_2$  (see Equation 3.20) (D) Variability in gene expression as measured by the standard deviation of all cell states  $\sigma(X_k)$ . (E) Variability of the populations of identified ‘ON’ and ‘OFF’ cells.

essential information required to explain the persistence of bistability across a range of  $a_0$  values. The main idea of the method is to reduce the full system with  $N$  interacting cells to an description involving only a single cell coupled to an effective medium representing the rest of the system Fig. 3.7B. This is analogous to mean-field theory in statistical physics, where for the coupling of a one element to the rest of the system, one replaces a sum over all elements by a single population-averaged term. However, our goal here was not to use the mean-field description to find an approximate solution to our system, but to deduce statements on the range of possible steady state solutions our system can have.

In the mean-field picture, we first choose a random cell  $k$  and rewrite the signaling molecule contribution from the rest of the system as

$$\begin{aligned} Y_k^{other} &\equiv \sum_{l \neq k} f(r_{kl}) C(X_l) = \sum_{l \neq k} f(r_{kl}) ((C_{ON} - 1)X_l + 1) \\ &= (C_{ON} - 1) \sum_{l \neq k} f(r_{kl}) X_l + \sum_{l \neq k} f(r_{kl}) \\ &= f_N \left( (C_{ON} - 1) \frac{1}{f_N} \sum_{l \neq k} f(r_{kl}) X_l + 1 \right) \equiv f_N \bar{C}(X_{MF}). \end{aligned} \quad (3.21)$$

In the last step we introduced  $X_{MF} \equiv \frac{1}{f_N} \sum_{l \neq k} f_{kl} X_l$  as the effective mean-field state of the environment of cell  $k$ . By construction, the mean-field variable  $X_{MF}$  can take any value on  $[0, 1]$ , since all  $X_l$  are continuous on  $[0, 1]$ . Therefore, in this reduced picture our system is reduced to a two-variable system with  $(X_{cell}, X_{MF})$  on  $[0, 1] \times [0, 1]$ , where we write  $X_{cell} = X_k$ . The time-evolution of the single cell can then be written as

$$\begin{aligned} X_{cell}(t+1) &= \frac{Y_{cell}(t)^n}{K^n + Y_{cell}(t)^n} = \frac{(C(X_{self}(t)) + f_N C(X_{MF}(t)))^n}{K^n + (C(X_{self}(t)) + f_N C(X_{MF}(t)))^n} \\ &\equiv f_{MF}(X_{cell}(t), X_{MF}(t)). \end{aligned} \quad (3.22)$$

Note that up till this point, our description is exact, i.e. the time-evolution equation 3.22 describes the exact evolution of cell  $k$ . To obtain the dynamics of the reduced two-variable system, we would need to derive a time evolution equation for  $X_{MF}$ , which we did not attempt here. Instead, we focused on the steady states of the reduced system considered as a single-variable system for  $X_{self}$  with  $X_{MF}$  as a constant parameter. This is because a steady state of the full system (Eqs. 3.10, 3.12) is necessarily a steady state of this reduced mean-field system. Namely, suppose that  $(X_1, \dots, X_N)$  is a steady state of our full model, such that  $X_k(t+1) = X_k(t)$  for all  $1 \leq k \leq N$ . For any cell  $k$ , we can then calculate a mean-field variable  $X_{MF}^{(k)}(t) \equiv \frac{1}{f_N} \sum_{l \neq k} f_{kl} X_l(t)$ , such that

the update equation Eq. 3.22 with  $f_{MF}(X_k(t), X_{MF}^{(k)}(t))$  gives the same result for  $X_k(t+1)$  as Eq. 3.12. Therefore,  $f_{MF}(X_k(t), X_{MF}(t)) = X_k(t)$  and  $X_k(t)$  is a steady state of the mean-field model. However, the converse does not need to be true: a steady state of the mean-field model does not necessarily correspond to a steady state of the full system. Suppose we have fixed point of the reduced system 3.22. This corresponds to a set of variables  $(X_{self}, X_{MF})$  such that  $X_{self}$  remains constant under Eq. 3.22. However, since for the full system to be in equilibrium we must have a fixed point for every cell that we choose to be  $X_{self}$ . Therefore, this corresponds to finding  $N$  fixed points of the form  $(X_k, X_{MF}^{(k)})$  such that  $f_{MF}(X_k, X_{MF}^{(k)}) = 0$ . Since  $X_{MF}^{(k)}$  is a function of the  $N-1$  variables  $X_1, \dots, X_{k-1}, X_{k+1}, \dots, X_N$ , there are only  $N$  variables for the  $N$  constraints. Hence, there is no guarantee that there is a solution to this set of nonlinear equations. Thus, any fixed point of our full model must be a steady state solution for  $X_{self}$  in Eq. 3.22, but not every steady state of Eq. 3.22 corresponds to a fixed points of the full model.

We computed the fixed points of the mean-field model by numerically solving for  $X_{self}$  by setting  $X_{self} = f(X_{self}, X_{MF})$  in Eq. 3.22, while treating  $X_{MF}$  as a continuous parameter that we varied. This results in a cubic equation in  $X_{self}$ , which implies that for any fixed value of  $X_{MF}$ , there is at least one real and at most three solutions for  $X_{self}$ . The resulting solutions are continuous curves in the two-dimensional space spanned by  $X_{self}$  and  $X_{MF}$  space (Fig. 3.7C). Note that the intersection  $X_{self} = X_{MF}$  precisely corresponds to the uniform lattice solutions, since they coincide with the solutions of Eq. 3.16.

As  $a_0$  is varied from the synchronous regime to the autonomous regime, we observe the appearance of an extra solution branch consisting of an a low-valued stable solution and intermediate unstable solution (middle plot with  $a_0 = 4.0$  in Fig. 3.7C), which is present for a set of low values for  $X_{MF}$ . This means that a cell in a weakly secreting environment can be bistable, but if its surrounding concentration of signaling molecules exceeds some threshold (determined by the location of the saddle-node bifurcation in the mean-field solution plot), then it is forced to the high steady state. As  $a_0$  is further increased, the lower branch moves to the right, until the bifurcation point reaches  $X_{MF} = 1$ , after which the cell state always has three steady states regardless of the environment (right plot with  $a_0 = 7.0$  in Fig. 3.7C).

We furthermore observed that the solutions varied less with  $X_{MF}$  as  $a_0$  was increased. This is intuitive, since a weaker interaction of the cells implies that the steady states depend more on the internal dynamics of a cell (due to self-communication) and less on its external coupling to the environment. We can quantify this dependence on  $X_{MF}$  by considering the spread in values for each solution branch of the mean-field model. We measured this by taking the difference between the maximum and minimum values of that branch, where for the lower branch (or whichever has two solutions) we take only

the stable solution into account. Formally, let us denote  $X_{SS}^{(i)}(X_{MF})$  as the steady states solution of  $X_{self}$  as a function of  $X_{MF}$ , and  $i = \text{“ON”}, \text{“OFF”}$  denote the stable solutions of the system. Let  $D^{(i)} \subseteq [0, 1]$  denote the domain of  $X_{SS}^{(i)}$  on which there is a solution. Then for each solution we can define the maximal spread in values due to environmental changes as

$$\sigma_{\max}^{(i)} \equiv \max_{X_{MF} \in D^{(i)}} \left( X_{SS}^{(i)}(X_{MF}) \right) - \min_{X_{MF} \in D^{(i)}} \left( X_{SS}^{(i)}(X_{MF}) \right). \quad (3.23)$$

This quantity  $\sigma_{\max}^{(i)}$  is also graphically depicted in the steady state solution plots in Fig. 3.7C.

We then examined how this maximal variability changed as we tuned the value of  $a_0$ . As expected,  $\sigma_{\max}$  generally decreased as a function of  $a_0$ , with some difference between the “OFF” and the “ON” cells (left plot of Fig. 3.7D). The curve for the “OFF” cells starts around  $a_0 \approx 2.8$  since this is where the lower solution branch first appeared. It then first increases in value, but from about  $a_0 \approx 6$  onward both maximal variability values for “ON” and “OFF” cells decrease with  $a_0$ .

Furthermore, we verified that these theoretically derived upper bounds are consistent with the variability in the steady states found in simulations (right plot of Fig. 3.7D). We performed multiple simulations for each  $a_0$  and determined the variability from the standard deviation in cell states for both the “ON” and the “OFF” cells. We found for all simulations that the actual variability was indeed lower than the theoretical upper bounds determined. On average, the simulated variabilities were considerably lower than the theoretical upper bounds, with a higher variability for the “OFF” cells than for the “ON” cells, in accordance with expectations.

### 3.2.6. DISCUSSION

In this section, we introduced and analyzed an extension of the multicellular model introduced in Section 3.1, in which the cells have a continuous secretion rates and respond to sensed concentrations of signaling molecules through a sigmoidal response function. We identified two distinct qualitative phases of this model, which occur when the cell density is very high or very low. At high density, the cells strongly couple to each other and tend to synchronize their gene expression levels, such that the steady state is always a homogeneous lattice where all cells have identical states. The steady states of this system can be found by solving a one-dimensional system for a homogeneous lattice, which is equivalent to that of a single cell up to rescaling of parameters. At low density, the coupling between cells is weak and the cells are autonomous, responding mostly to their own signaling molecules. The dynamics in this regime is qualitatively identical to that of a system of  $N$  independent cells. The boundary between these two regimes is not sharp and there exists a large transition region in which less simplistic dynamics occurs. Here,

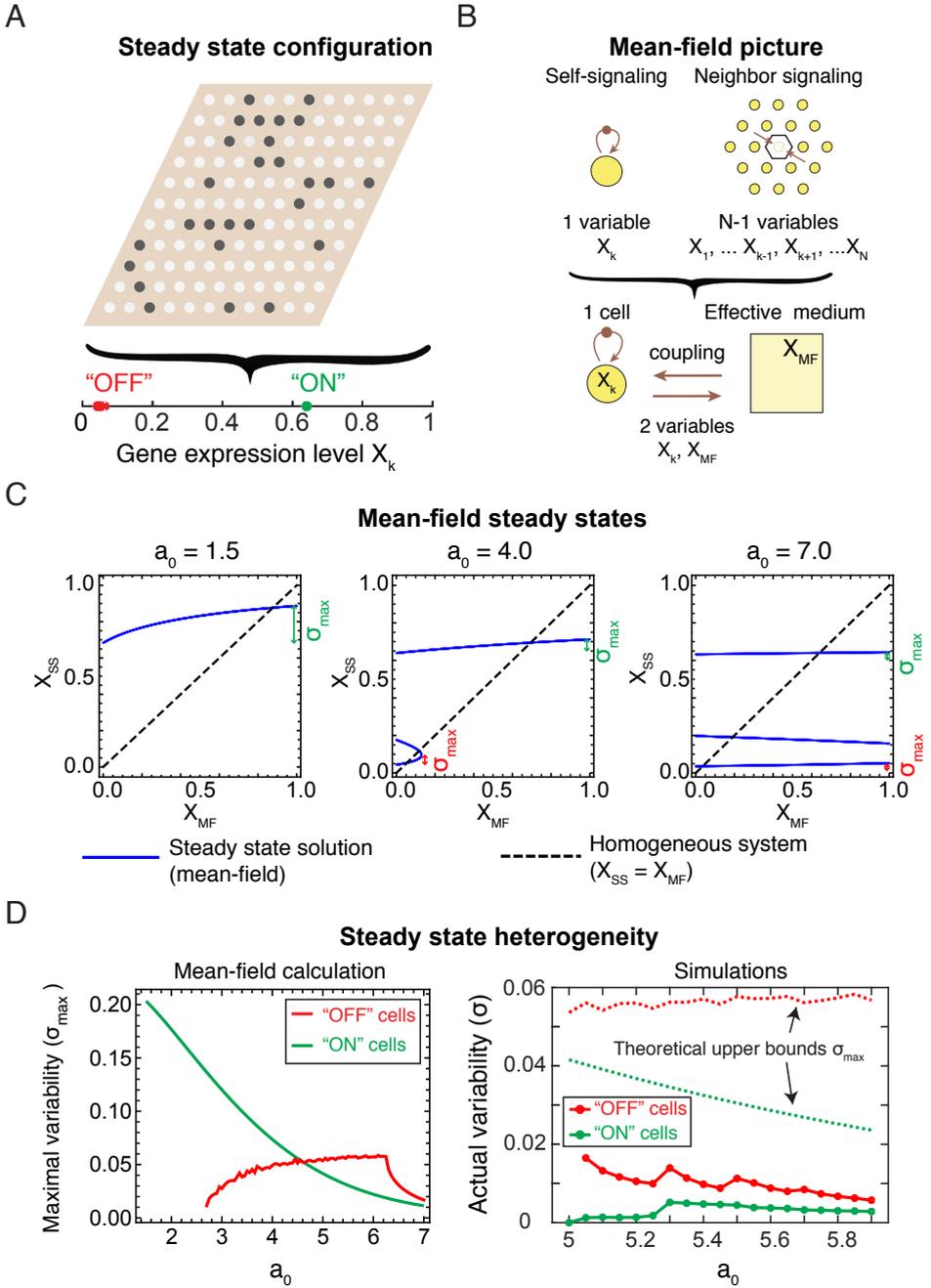


Figure 3.7: Mean-field analysis of steady states. (Caption continued on next page)

**Figure 3.7 (previous page):** (A) Example of a steady state configuration of the system. Darker colors correspond to higher levels of gene expression. The final states of the cells are mapped onto the one-dimensional axis and are observed to fall within two tightly bound clusters, which we identify as “ON” and “OFF” cells. (B) Schematic of the mean-field picture. The dynamics of the full system is described by  $N$  variables. In the mean-field approach, we select a cell  $k$  and map the  $N - 1$  variables describing the rest of the system onto a single variable describing the effective cell it is in. The two-variable system can be solved and its solutions are easy to visualize. (C) Solutions of the mean-field model for various interaction strengths (described by the inter-cell spacing  $a_0$ ). At low  $a_0$ , there is a single solution, but its value can vary considerably depending on  $X_{MF}$ . The spread in the values is denoted  $\sigma_{max}$  and is a measure for the maximum variability in gene expression between individual cells. At higher  $a_0$ , two more solutions appear in the form of a saddle-node bifurcation. The upper branch (in between the two other solutions) is unstable while the lower branch is stable. As  $a_0$  is further increased, this solution spreads across a larger range of  $X_{MF}$  values until it percolates across the entire range of  $X_{MF}$ . (D) The maximum spread in the mean-field solution is an upper bound for the variability between cells of the same solution branch (i.e., among “OFF” cells and “ON” cells). Left: theoretical calculations for  $\sigma_{max}$  across a range of  $a_0$ . Right: the actual variability in the two clusters of cells (“ON” and “OFF”) in simulation steady states. The dotted lines indicate the calculated theoretical upper bounds. For each  $a_0$ , we performed 500 simulations and calculated the standard deviation of cell states among “ON” and “OFF” cells for each steady state. The reported values are averages over these 500 simulations.

the system may partially activate or deactivate and form patterned islands of cells with identical gene expression levels. To characterize the transition between the autonomous phase and the collective phase, we introduced and discussed various statistical metrics that characterize certain aspects of the transition. The fraction of non-uniform final states, the fraction of “ON” cells, the Hamming distance between initial state and final state and a modified spatial index all allow us pinpoint the exact boundary of one of the two phases with the transition region, and appear to vary smoothly as the distance between neighboring cells is changed. Remarkably, with suitably chosen parameters, the system retains a binary character throughout all phases, whereby the steady state cells separate into a group of “ON” cells and a group of “OFF” cells that have very close gene expression levels. This phenomenon can be explained through a mean-field picture, where we reduce the system to a single cell coupled to an effective environment. Solutions of the mean-field model show the sensitivity of the single-cell steady states to the environment of the cell.

The synchronization of cell states when the cells interact strongly is qualitatively not surprising, but begs the question whether such behavior arises from more general mathematical principles underpinning our model. Such principles could be searched for in the literature on coupled map lattices, which are models of discrete coupled units with continuous state variables that evolve in discrete time [Kaneko, 1993]. The collective tuning of gene expression levels is also highly reminiscent of synchronization of coupled identical oscillators. Such systems have seen an overwhelming number of studies in the context of mechanical oscillators, biological clocks, coupled neurons and many other linked oscillators [Strogatz, 2004]. In our system, changing the response function from a sigmoid with positive slope (positive feedback, activation) to a sigmoid with negative

slope (negative feedback, repression) produces a system that shows oscillations in gene expression level. At high density, the oscillations of different cells synchronize while at low density individual cells oscillate independently. Further work should elucidate the connection between our model and more established models of coupled oscillators.

One particular approach that could be relevant for our system is the derivation of dynamical equations for the moments of the system. Here, one first defines a number of statistical quantities of a system of many coupled units such as the mean, variance and higher moments of some variable one is interested in. One can then often derive exact mathematical expressions for the time evolution of these moments, which match well with simulations of the original, possibly stochastic model. This approach has proven useful for a variety of systems of globally coupled units including coupled oscillators [De Monte & d'Ovidio, 2002; De Monte et al., 2003], ecology [Bolker et al., 1997], chemotaxis [Young et al., 2019] and chemical reaction networks [Gonzales et al., 2019]. In our system, one particular result that we wish to derive is to show that the variance of the cell states goes to zero in the strong interacting regime, whereas it remains positive in the weakly interacting regime.

Finally, there are many features of biological systems that could be added to make the model more realistic and more widely applicable. Several such extensions have been defined and studied in the extended model involving multiple genes, which will be discussed in Chapter 5, but could also be included in the present model. Heterogeneity between cells could be implemented by randomly generating different parameters for each cell. Stochastic dynamics due to stochastic gene expression, limits to sensing precision and other random effects should be included. Different spatial arrangements of cells in both lattice and non-lattice settings should be studied. Effects such as cell motility, mechanical interactions and hydrodynamic flows could also be integrated into the model.

### 3.2.7. APPENDIX

For a uniform lattice or a single cell, we can exactly solve for the steady states in a number of special cases. Here we consider the cases  $n = 1$  and  $n = 2$  and analytically derive their steady state solutions.

**Michaelis-Menten  $n = 1$  system is monostable** For non-ultrasensitive sensing, corresponding to a Hill coefficient  $n = 1$ , the system is always monostable. This is evident from plotting  $X_m(t + 1)$  against  $X_m(t)$ , but also follows from a straightforward calculation. The steady states are obtained from Eq. 3.16 with  $f_m(X_m) = X_m$ , which for  $n = 1$

becomes a quadratic equation of the form  $aX_m^2 + bX_m + c = 0$  with

$$\begin{aligned} a &= (C_{ON} - 1)(1 + f_N) \\ b &= (C_{ON} - 2)(1 + f_N) - K \\ c &= -(1 + f_N) \\ \Delta &= b^2 - 4ac = (K - C_{ON}(1 + f_N))^2 + 4(1 + f_N)K. \end{aligned} \quad (3.24)$$

Since  $\Delta > 0$  for any choice of parameters, there are two real solutions to the quadratic equation. However, because  $a > 0, c < 0$ , we have  $\Delta > b^2$ . Therefore, one of the solutions is negative. It follows that there is only one positive solution given by  $X_m^* = \frac{-b + \sqrt{\Delta}}{2a}$ . Note that by construction of Eq. 3.16,  $0 \leq X_m^* \leq 1$ .

**Ultrasensitive  $n = 2$  system has a region of bistability** With Hill coefficient  $n = 2$ , the solution to  $X_m(t + 1) = X_m(t)$  (Eq. 3.16) gives a cubic equation in the parameters  $a_0, K$  and  $C_{ON}$ . This implies that there is at least one and at most three fixed points for the uniform system with  $n = 2$ . The solutions of this cubic equation are determined by the determinant

$$\begin{aligned} \Delta &= 18abcd - 4b^3d + b^2c^2 - 4ac^3 - 27a^2d^2 \\ a &= -(C_{ON} - 1)^2 \\ b &= (C_{ON} - 1)^2 - 2(C_{ON} - 1) \\ c &= \frac{2(C_{ON} - 1) - K^2 - 1}{(1 + f_N)^2} \\ d &= 1. \end{aligned}$$

If  $\Delta > 0$ , there are three real solutions, two stable fixed points with an unstable fixed point in between. If  $\Delta < 0$ , we have only one stable fixed point. The border between the bistable and monostable regions is determined by  $\Delta = 0$ . This expression can be reduced to a trivial solution  $K^2 = 0$  and an expression that is quadratic in  $K^2$ ,

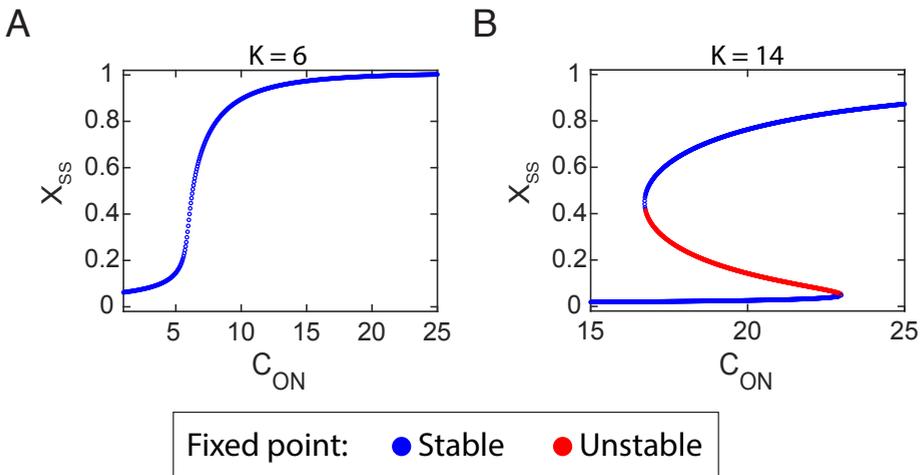
$$-4C_{ON}^3 + \frac{(C_{ON}^2 + 18C_{ON} - 27)}{(f_N + 1)^2}K^2 - \frac{4}{(f_N + 1)^4}K^4 = 0. \quad (3.25)$$

This quadratic equation in  $K^2$  can be solved straightforwardly, and since  $K$  is positive we directly obtain solutions for  $K$  as a function of  $C_{ON}$  and  $f_N$ . Depending on the values of these parameters, there can be either two or no solutions to this equation. In the case of two solutions, these correspond to the upper bound and lower bound of bistable region.

**Stability of the uniform lattice solution ( $n = 2$ )** For the one-dimensional model of a uniform lattice (Eq. 3.16), stability is easily determined. Whenever the system has three fixed points, the outer two are stable and the middle one is unstable. However, we would like to know whether the stable fixed points are also stable in the full system where cells can attain different expression levels. In other words, we would like to know whether the system is stable with respect to perturbations in the larger phase space  $[0, 1]^N$ . This can be done by taking the linearization of Eq. 3.12, which we will denote as  $f(X(t)) \equiv X(t+1)$  for the moment, and computing eigenvalues of the Jacobian matrix  $Df(X)$  evaluated at the fixed point  $X_{SS}$ :

$$[Df(X_{SS})]_{ij} = \left. \frac{\partial f^{(i)}(X)}{\partial X_j} \right|_{X=X_{SS}} \quad (3.26)$$

The system is stable if and only if all eigenvalues of  $Df(X_{SS})$  have norm smaller than one (i.e.  $|\lambda_i| < 1$  for all  $1 \leq i \leq N$ ). Using this approach, we verified that the uniform lattice fixed points show the same stability as the one-dimensional system (bifurcation diagrams of solutions at constant  $K$  are shown in Fig. 3.8). The system is either monostable with a single fixed point, or bistable with an intermediate unstable fixed point. Unsurprisingly, the same stability behavior persists in the autonomous phase (not shown). Here, the system approximates the uncoupled case (without cell-cell interaction), where the fixed points that are stable in the one-dimensional model are trivially stable in the  $N$ -dimensional system.



**Figure 3.8: Stability of the uniform lattice solution for the bistable system with  $n = 2$ .** (A-B) Bifurcation diagrams with steady state solutions for  $X_m$  as a function of  $C_{ON}$ , at fixed values of  $K$ . Stable solutions are shown in blue, unstable solutions in red. Stability is determined by computing largest eigenvalue  $\lambda_{\max}$  of the linearized map of Eq. 3.12 around each of the fixed points in the system. Parameters:  $N = 121$ ,  $a_0 = 1.5$  (weak interaction), (A)  $K = 6$  (monostable), (B)  $K = 14$  (intersects bistable region).

### 3.3. SENSITIVITY TO INITIAL CONDITIONS

One key question in complex systems research is how local perturbations spread to affect the collective behavior of a system. In multicellular systems, one method of experimentally applying localized perturbations is through shining light on engineered strains of microorganisms to induce gene expression changes – these techniques were first applied to neurons and have given rise to the field of optogenetics [Deisseroth, 2013]. Concretely, one can engineer cells with proteins that only get activated in sufficiently light conditions and shine a localized beam of (laser) light onto single cells to activate gene expression. In the social amoeba *Dictyostelium discoideum*, manipulating a single cell can cause an entire population of starved cells to move towards that cell and aggregate [Sgro, 2019]. From a theoretical point of view, sensitivity of initial conditions has been widely studied in chaotic systems and is one of the hallmarks of chaos. For discrete-time systems, various mathematical methods have been developed to formally analyze sensitivity to initial conditions in cellular automata ([Wolfram, 1983; Wolfram, 1984; Urías et al., 1997]) and coupled map lattices ([Kaneko, 1986]). These methods typically rely on defining new measures and statistical quantities to quantify the effect of the perturbation. We adapted one such metrics (the Hamming distance) and constructed new ones to quantify both the strength of the perturbation and the response. In the following, we first consider digital cells (Section 3.1), and then analogue cells (Section 3.2).

#### 3.3.1. DISCRETE SYSTEM (INFINITE HILL COEFFICIENT)

**Applying perturbations and quantifying their effects** Given an initial configuration, we generated configurations which are closely related to it by flipping the state of a small number of cells. For instance, given a configuration  $X = \{X_k\}_{k=1}^N$ , with  $X_k = 1$  corresponding to the ON state and  $X_k = -1$  corresponding to the OFF state, choose a random cell  $k \in \{1, 2, \dots, N\}$  and flip it through the transformation  $X_k \rightarrow -X_k$  (Fig. 3.9A). The trajectories that follow from this operation typically closely follow each other (Fig. 3.9B), but nevertheless show a variation in both trajectory shapes and endpoints. For instance, there is a metastable state that terminates around  $p \approx 0.6$  in Fig. 3.9A while all other trajectories terminate at  $p = 1$ . We examined how sensitive our system is to slight changes in initial conditions using different ways of perturbing the system and measuring its effect. The impact of the perturbation was measured through the difference in equilibrium states between the unperturbed and the perturbed trajectory. This difference can be quantified through the Hamming distance introduced in Section 3.2.4 (Eq. 3.19). More precisely, let  $X_{eq}$  be the final state of the unperturbed trajectory and  $Y_{eq}$  be the final state

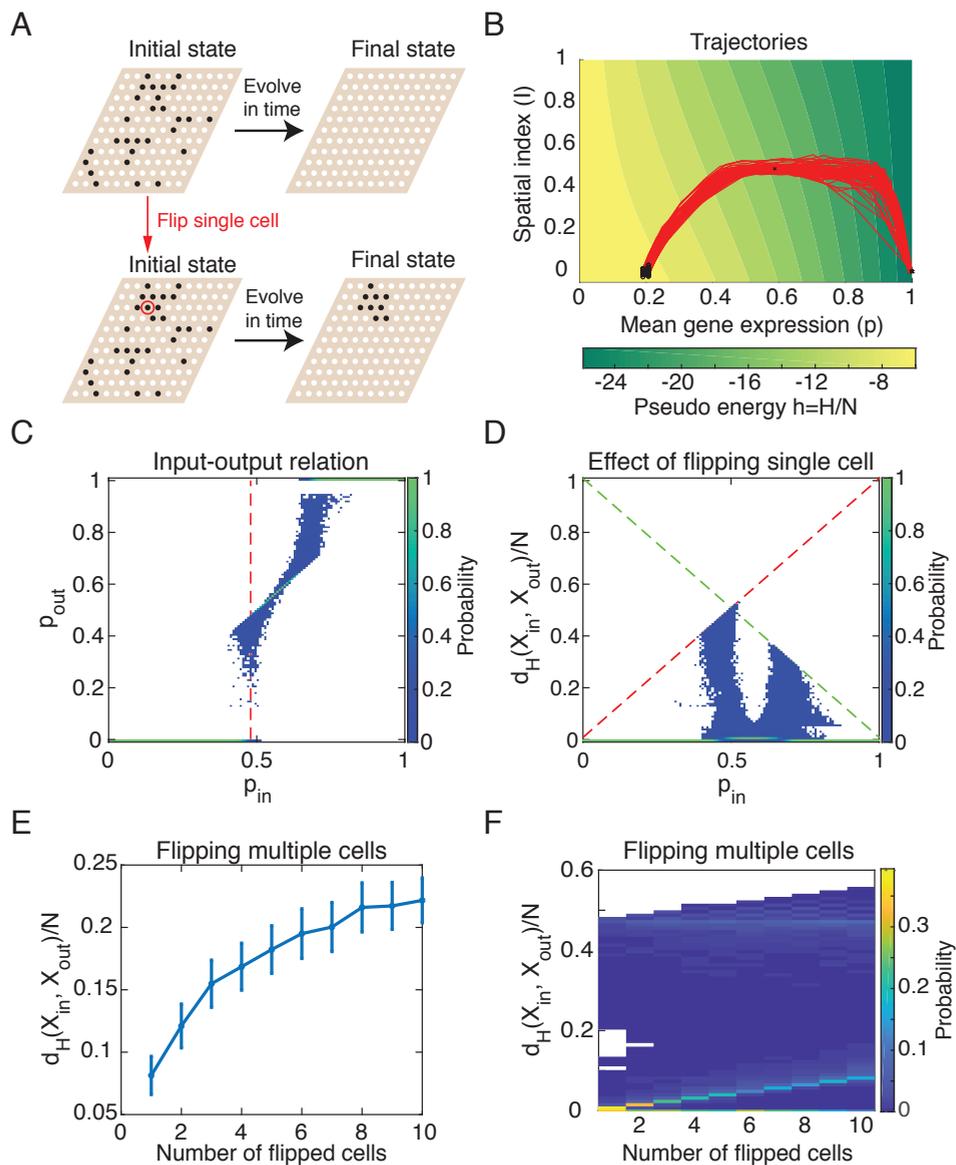
of the perturbed trajectory. Then we quantified their difference through

$$d_H(X, Y) = \sum_{k=1}^N |X_k - Y_k|, \quad (3.27)$$

Recall that biologically, this quantity sums up the differences in gene expression level of all the individual cells.

**Flipping single cells** Using this metric, we examined how far apart the final states are for two trajectories with close initial states. We first studied the effect of flipping single cells across different phases of the system. Instead of varying many parameters, we fixed most parameters at values corresponding to the activation-deactivation phase, which shows behavior from all different qualitative phases found in the system as one varies the initial fraction of ON cells. As seen from the  $p_{in} - p_{out}$  map in Fig. 3.9(C), the system shows dynamics characteristic of all five phases (all OFF, deactivation, autonomy, activation and all ON) as  $p_{in}$  is increased from 0 to 1. Therefore, we studied this single example and extrapolated our results to other parameter sets not examined here.

As expected, the system is most sensitive to changes in initial conditions in the ‘active’ phases (activation and deactivation), where it has many possible final states with different values of  $p_{out}$  (Fig. 3.9D). In the other phases, the response is not responsive to perturbations if these are sufficiently small. In the all ON/OFF regions flipping a single cell has no effect, because it produces another configuration where all the cells turn ON or OFF. In the autonomous region, any change to the system results in a new steady state configuration, so if we change a single cell, the new equilibrium configuration will differ from the old by only a single cell. However, in the activation and deactivation regions, a small change can have a large effect. In particular, if we take a  $p_{in}$  value at which some configurations partially turn ON/OFF whereas others completely turn ON/OFF, then changing a single cell might provide the ‘kick’ necessary to fully activate or deactivate the system. Hence a metastable state consisting of an island with a few ON cells could turn OFF all these cells if we change just a single cell. The trajectories that show this behavior are marked by the locations of the red dotted lines in (Fig. 3.9D). These represent the maximum number of cells that can be turned OFF (red line) or ON (green line) for any value of  $p_{in}$ . For deactivation,  $d_H/N$  is bound by  $p_{in} + 1/N$ , since no more than  $p_{in}N + 1$  cells can turn off (the  $p_{in}N$  cells which are ON initially plus the cell that is flipped). For activation,  $d_H/N$  is bound by  $1 - p_{in} + 1/N$ . The fact that there are trajectories intersecting with this line shows that full (de)activation due to changing a single cell can indeed happen in the system, for a variety of initial conditions.



**Flipping multiple cells** We next examined the effect of changing the state of multiple cells of the initial state on the final configurations. To do this, we used the same approach as before and extended it to arbitrary number of cell flips. The average distance between the flipped configuration and the original configuration is found to increase

with the number of flipped cells, as expected. The increase saturates as the number of cells increases (Fig. 3.9E), since  $d_H/N \leq 1$  is a strict upper bound. In practice, the bound is roughly  $\max(p_{in}, 1 - p_{in})$  of the initial state of the perturbed system, since at best all the ON (OFF) cells turn OFF (ON), but we never observed activation and deactivation happening simultaneously in one trajectory. Furthermore, from the probability distributions in Fig. 3.9F we see two peaks for each of the different numbers of flipped cells. The first one is where  $d_H$  equals the number of flipped cells and corresponds to trajectories where the perturbed and unperturbed configuration were both autonomous (i.e. their final states are their initial states), so that their final states differ only by the cells that were flipped. The second peak occurs at around  $d_H \approx 0.48N = p_{in}N$  and corresponds to cases where one of one of the trajectories turns OFF completely while the other one remains autonomous. For instance, the unperturbed configuration could be a steady state, while perturbing this state leads to a configuration where all of the  $p_{in}N$  ON-cells turn OFF. We concluded that perturbing cells in an active phase (deactivation in this case) tends to be dominated by two extreme effects. Either the system is completely unresponsive to the perturbation, leaving only the perturbed cells affected, or the perturbation causes the system to escape the metastable state it was in, thereby traveling further down the pseudo-energy landscape towards one of the global minima consisting of all cells ON or OFF.

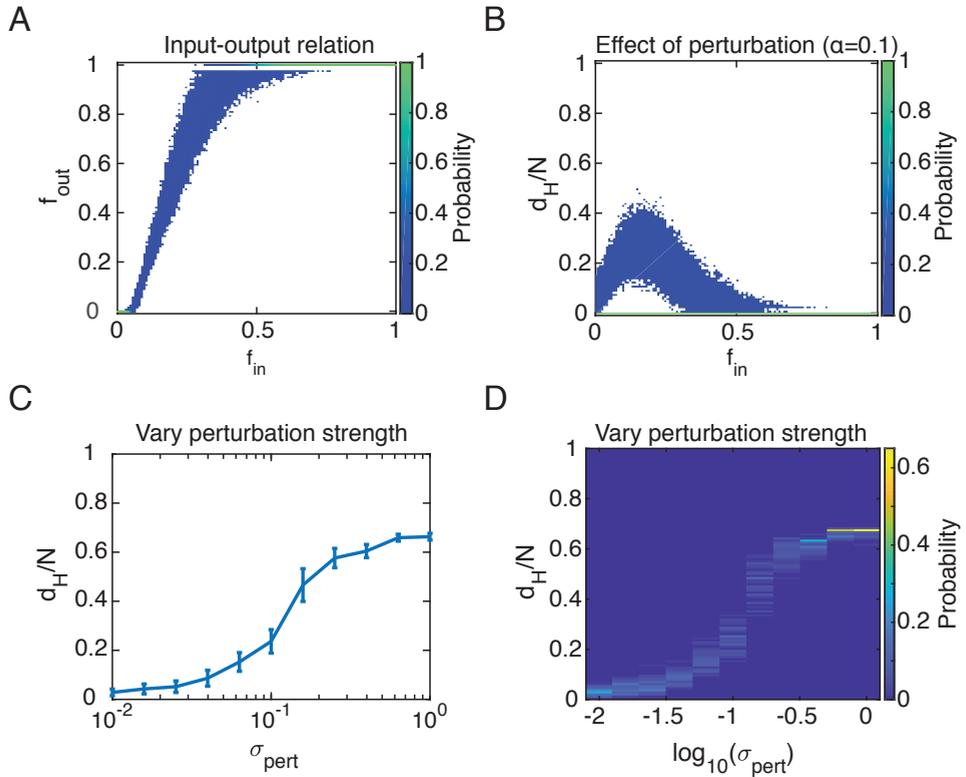
### 3.3.2. CONTINUOUS SYSTEM (FINITE HILL COEFFICIENT)

Let us now consider the case where cells have a continuous state and continuous response functions as described in Section 3.2.1. Since we do not have binary cells which are either ON or OFF, the method of flipping cell states cannot be applied to generate

---

**Figure 3.9 (previous page): Perturbation response of a population of digital cells.** (A) Schematic of the perturbation method. (B) Configurations that differ by a single cell are likely to end up in the same final states. Red lines are trajectories starting from the circles and ending at the crosses. The background color indicates the value of the pseudo-energy (see ...).  $N = 121, a_0 = 1.5, K = 6, C_{ON} = 21, p_{in} = 0.2$  (Activation phase, weak interaction). (C-D) Effect of flipping a single cell for various initial  $p_{ini}$ . (C) Relation between the initial fraction of ON-cells  $p_{in}$  and the final fraction of ON-cells  $p_{out}$ . Results are for a system in the activation-deactivation phase. The dotted red line shows the  $p_{in}$  value taken in plots (E) and (F) (see below). (D) The average distance between final configurations of trajectories as a single cell's state is changed. We took the trajectories that were used to generate (B). For each trajectory, we flipped the state of a one single cell and measured the distance between the final configurations of the flipped and non-flipped states. The reported values are averages over the  $N$  possibilities of choosing a single cell. Activation-deactivation phase (weak interaction).  $N = 121, a_0 = 0.5, K = 10, C_{ON} = 5$ . (E-F) Effect of flipping multiple cells. The initial configuration for all data is  $N = 121, a_0 = 0.5, K = 10, C_{ON} = 5, p_{ini} = 58/N \approx 0.48$  (the dotted red line shown in (C)). (E) Distance between final configurations plotted against number of cells of the initial configuration we flipped. Error bars are standard errors of the mean. (F) Probability densities of the distance between final configurations for multiple numbers of flipped cells. Each row shows  $P(d_H/N)$  (see color bar) for a given number of flipped cells. The white regions have probability 0.

perturbations to the system. In the continuous model, the range of possible perturbations and responses is more complicated, and required us to define new ways to perturb the system and analyze its effects.



**Figure 3.10: Perturbation response of a population of analogue cells.** We work in the binary description of the continuous cells (see text). All results are for  $N = 121$ ,  $a_0 = 5.4$ ,  $K = 8$ ,  $C_{ON} = 16$ , which corresponds to the transition region between the synchronous and the autonomous phase. (A) Relation between the initial fraction of ON-cells  $f_{in}$  and the final fraction of ON-cells  $f_{out}$ , where ‘ON’ and ‘OFF’ are defined for continuous cells according to Eq. 3.18 in Section 3.2.4. (B) Distance between perturbed and unperturbed trajectories. The perturbation strength is set at  $\alpha = 0.1$ . (C-D) Effect of perturbation strength on variability in the final states of trajectories, for trajectories with  $f_{in} = 0.3$ . (C) Distance between final configurations plotted against perturbation strength. Error bars are standard deviations from the set of different perturbed trajectories. (D) Probability densities of the distance between final states for various levels of perturbation strength.

**Generating initial states** There are various ways to generate initial states for the continuous system, and these can lead to considerably different dynamics and steady states. Because a homogeneous configuration (where all the cells have the same state) remains homogeneous and corresponds to a one-dimensional system, we did not consider this case here. Instead, we used a two-step algorithm to generate initial states with controllable properties outlined below.

First, we fixed the initial fraction of cells  $f_{ini}$  which are ‘ON’ according to the binary description we constructed (Eq. 3.18 in Section 3.2.4). We initiated a random configuration with  $Nf_{ini}$  ‘ON’ cells and  $N(1 - f_{ini})$  ‘OFF’ cells. This is analogous to the discrete system with infinite Hill coefficient, except that the cells have gene expression levels at the locations of the stable fixed points  $X_{OFF}$  and  $X_{ON}$  with  $0 < X_{OFF} < X_{ON} < 1$ , obtained by solving the system for a single isolated cell. In the second step, we added variability to the gene expression levels using a Monte Carlo algorithm that keeps  $f_{in}$  constant. We selected random cells  $k, l$ , and generated a noise term  $\delta X \sim \mathcal{N}(0, \sigma)$ , with  $\sigma$  sufficiently small. We then attempted to set  $X_k \rightarrow X_k + \delta X$  and  $X_l \rightarrow X_l - \delta X$ . If the resulting values are such that ‘ON’ cells remain ‘ON’ and ‘OFF’ cells remain off (and if the values are in  $[0, 1]$ ), then the move is accepted. Otherwise, the system does another trial with a new  $\delta X$ . Note that this change keeps both the fraction of ‘ON’ cells  $f_{in}$  and the average gene expression  $p_{in}$  constant. As the number of iterations is increased, the variance of the cell states  $\text{Var}(X_i) = \sum_{k=1}^N (X_k - \langle X_k \rangle)^2$  plateaus, since the cells can only differ from each other by a finite amount. In practice, we always worked with a high number of iterations in the saturating regime.

**Perturbation strength** We defined a perturbation strength that measures how strongly we altered the initial state of the system. We perturbed the initial state generated using the same procedure as above, by adding random noise terms (which do not necessarily preserve the mean expression level). For each cell, we tried to perturb its state by adding a noise term  $\delta X_k \sim N(0, \sigma_{pert})$ . Assume that the noise terms of different cells are uncorrelated. If  $X_k + \delta X_k \in [0, 1]$ , the move is accepted and we set  $X_k \rightarrow X_k + \delta X_k$ . Otherwise, the system tries a new  $\delta X_k$  and repeats this until the move is accepted.

Using the same definition for  $d_H$ , the initial distance between the perturbed and unperturbed states is then

$$d_H(t=0) = \sum_{k=1}^N |\delta X_k|. \quad (3.28)$$

The mean of this quantity, which is the average initial distance, is then

$$\begin{aligned} \langle d_H(t=0) \rangle &= \sum_{k=1}^N \langle |\delta X_k| \rangle = 2N \int_0^\infty \frac{x}{\sqrt{2\pi}\sigma_{pert}} \exp\left(-\frac{x^2}{2\sigma_{pert}^2}\right) dx = \sqrt{\frac{2}{\pi}} N \sigma_{pert} \\ &\approx N \sigma_{pert}. \end{aligned} \quad (3.29)$$

On the other hand, the effect on the mean expression level averages out to zero on average. If we denote  $p_0$  as the mean expression of the unperturbed state and  $p_1$  as that of

the perturbed state, we have

$$p_1 - p_0 = \sum_{k=1}^N \delta X_k, \quad \langle p_1 - p_0 \rangle = 0, \quad \langle (p_1 - p_0)^2 \rangle = \sigma_{pert}^2. \quad (3.30)$$

Thus, the initial mean expression level gets perturbed by the addition of a Gaussian noise term with mean 0 and standard deviation  $\sigma_{pert}$ .

### 3

**Dependence on initial fraction of ON cells** We used the same distance function (Eq. 3.27) to measure distances between final states, applied to the output states of the perturbed and unperturbed final states in the binary description. Recall that the continuous model has a strong interaction phase, in which all lattices become uniform over time, and a weak interaction phase, in which each cell evolves according to its internal dynamics. Both of these regimes, perturbations have a trivial effect. In the strong interaction regime, the system homogenizes regardless of the initial state. Perturbations could at best take the system from one homogeneous steady state to another. In the weak interaction regime, the cells are autonomous and therefore such perturbations can only have an effect on the level of individual cells and without inducing any collective response. Therefore, only considered the transition phase between these two trivial phases in more detail.

The example considered in Fig. 3.10 has features analogous to the activation phase for binary cells. At very low  $f_{in}$ , the lattices becomes homogeneous at  $f_{out} = 0$ , but as  $f_{in}$  is increased, the values for  $f_{out} > f_{in}$ . From about  $f_{in} = 0.5$ , the system practically always reaches  $f_{out} = 1$ . Hence, the interesting features are found in the region  $0 < f_{in} < 0.5$ , where we could expect small changes to have considerable effect. Indeed, we found that a weak perturbation ( $\alpha = 0.1$ ) can cause a considerable effect, evident in the values of  $d_H$  taken around  $f_{in} = 0.1 - 0.2$ . However, we also observed that the probability of cause such a large response is in fact very low and that most trajectories remain unaffected at  $d_H = 0$ .

**Dependence on perturbation strength** Analogous to flipping multiple cells in the binary model, we then adjusted the noise strength and examined the effect on the variability of final configurations (as measured by  $d_H/N$ ). This tells us how the system responds as perturbation strength is increased. The Hamming distance between trajectories  $d_H/N$  increases with perturbation strength  $\sigma_{pert}$  and plateaus at high values of  $\sigma_{pert}$  as expected. However, in this case we see that the distribution of  $d_H/N$  is unimodal rather than bimodal. Small perturbations cannot cause a large response such as

an event where the all the cells turn ON (because for this example we took  $f_{in} = 0.3$ , this event would imply  $d_H/N = 0.7$ , but the probability of observing this is zero except at the highest values of  $\sigma_{pert}$  we took). This lack of ‘catastrophic events’ could have more to do with the way we set up the perturbation and chose the parameters of our system than be due to some inherent property of the continuous system. For instance, changing all cells slightly may result in different results from changing one cell by a large amount. Also, in the phases of the binary system where the system shows a less varied collection of output states for each input state (e.g. close to autonomy), we typically also expect the perturbation response to be weak. Nevertheless, the key point here is that we still observe that a small fraction of trajectories shows high sensitivity to initial conditions as is evident from Fig. 3.10B, causing two initially similar configurations to diverge over time to the maximal extent possible within the constraints of the system (as determined by the  $f_{in}$ - $f_{out}$  maps).

### 3.3.3. DISCUSSION

In this section, we studied our system’s sensitivity to initial conditions by applying a single perturbation to the initial state at  $t = 0$  and analyzing its effects. For binary cells, the response in the insulating phases (all ON/OFF; autonomy) is trivial, but in the active phases (activation/deactivation) there are different possible outcomes. The  $p_{in}$ - $p_{out}$  maps already show that the equilibrium  $p_{out}$  value can differ considerably between trajectories starting at the same value of  $p_{in}$ , which is a natural result of the fact that many microstates have the same value of  $p_{in}$ . However, from this observation alone it was not yet clear whether small changes can lead to these large differences in the final state. For instance, if we observe for some value of  $p_{in}$  that some trajectories remain autonomous whereas other completely turn off, can the trajectories that show each of these behaviors differ by only a single cell? Or are the autonomous trajectories in some sense still clustered, and would one need to apply a considerable perturbation involving many cells to see a difference in final state? From our results, it is clear that the first scenario applies: small changes can lead to drastically different behavior between closely related initial states. This is the case for the binary system, where flipping the state of a single cell of a steady state configuration can cause all cells to turn ON or to turn OFF. In the case of continuous cells, the effects are more subtle, but we also observe that small perturbations to the initial state can lead to the large deviations between the perturbed and unperturbed trajectories.

Our system’s sensitivity to initial conditions is shared with chaotic systems. However, this condition alone is not sufficient to claim that a system is chaotic. We have not formally established whether our system is chaotic or not; this requires proper mathematical definitions of e.g. the topology of our phase space. However, intuitively we would

not directly expect the current system to be chaotic, as we have not observed many of the other typical features of chaotic systems (e.g. period doubling, signs of strange attractors, etc.).

### 3.4. USING MUTUAL INFORMATION TO CHARACTERIZE PATTERN FORMATION

#### 3.4.1. PREDICTING AND INFERRING INITIAL AND FINAL STATES

Our analysis so far has focused on the forward problem of determining the dynamics and final state of simulation trajectories from a given parameter set and initial conditions. However, one can also imagine situations in which the backward problem is interesting: given a particular pattern, under what conditions could it have been formed? In other words, how much information does the final pattern provide about how the system started? The theoretical framework relevant for discussing and quantifying these ideas is information theory. Traditionally, information theory has been applied to study signal processing and communication, for instance to optimize the transfer of information across a noisy channel or to optimally compress data [Shannon, 1948; Cover & Joy, 2006]. Information theory has also found applications in developmental biology, neuroscience and biochemical networks [Tkačik & Walczak, 2011; Tostevin & Ten Wolde, 2009; Tostevin & Ten Wolde, 2010]. One well-studied example is pattern formation in the early *Drosophila* embryo. Here, cells along the anterior-posterior axis acquire different fates depending on concentrations of morphogen deposited by the mother at one end of the embryo. The maternal morphogens regulate various downstream genes, which in turn causes cells at different positions to develop differently. Information theory has been applied to precise measurements of the morphogen level and expression of the downstream genes to address the question of how precisely cells can tell their position along the anterior-posterior axis [Tkacik & Walczak, 2011; Gregor et al, 2007b]. As such, it directly tests Wolpert's idea of positional information, which proposes that genetically identical cells subject to different chemical environments (morphogen concentrations) can acquire different fates responsible for pattern formation [Wolpert, 1971].

In this section, we discuss applying information theory to quantify how much information the initial and final states of a multicellular system contain about each other. However, note that our example does not directly involve signaling or communication in natural settings. A multicellular system does not use its gene expression dynamics to communicate information about what it will do in the future, neither to itself nor to an external observer. Vice versa, it also does not try to inform any observer about its past state through its present gene expression pattern. Hence the idea of a communication

channel through which information is transmitted is irrelevant in our case. On the other hand, one could imagine applications in biological computing making use of multicellular systems to store and transmit information in more artificial settings [Amos, 2014; Macia et al., 2016]. For such applications, our approach may provide useful results concerning the information that can be encoded into a dynamically evolving multicellular system that is to be retrieved later.

We start by introducing and define mutual information between input and output states. We then discuss how mutual information is related to ‘entropy of population’, a quantity defined in earlier work that quantifies the degree of autonomy of the system. We argue that mutual information is relevant, due to the difficulty of directly inferring past states from the rules of our system. Next, we extend the definition of mutual information to our system with continuous cell states. Finally, we apply our framework to quantify the effect of noise on the system.

**Prediction and inference** Let us first discuss the statistical concepts of prediction and inference. Prediction here refers to our ability to forecast the final state of our system from knowing only an initial state. Inference is the reverse process of finding initial states corresponding to a given final pattern. We have seen that with fixed parameters, different initial conditions can lead to very different final states, but this depends heavily on the behavioral phase of the system. We have studied how different initial fractions of ON-cells  $p_{in}$  relate to different final values of ON-cells  $p_{out}$  by running many different simulations across the range of possible  $p_{in}$ . Essentially, in this process we aimed to reconstruct the conditional probability distribution  $P(p_{out}|p_{in})$ . However, once we know this distribution, we can also answer the question of which initial conditions are most likely for any given final configuration. Formally, this process of inference is described by Bayes’ rule:

$$P(p_{in} | p_{out}) = \frac{P(p_{out} | p_{in}) P(p_{in})}{P(p_{out})}. \quad (3.31)$$

Thus, we see that prediction and inference are two sides of the same coin. By constructing the  $p_{in} - p_{out}$  maps, we have therefore acquired the data needed to infer initial state from final states. However, our ability to infer the initial state depends on more than only the data that we have. If each initial state leads to a unique final state (as is the case in the autonomous phase), then we can infer the initial state with perfect certainty. However, if all initial states lead to the same final state (as is the case in the all ON/OFF phases), then we can infer nothing about the initial state from the final state. This leads to the question of how to quantify our ability to learn about the past (the initial state of the system) from the steady state that we observe at the end of a simulation.

**Mutual information between initial and final fraction of ON cells** Using information theory, we can quantify our ability to learn about one variable by studying another through the concept of mutual information. The mutual information between the two variables  $p_{in}$  and  $p_{out}$  is defined as

$$I(p_{out}, p_{in}) = S[P(p_{out})] - S[P(p_{out}|p_{in})], \quad (3.32)$$

where

$$S[P(p)] = - \sum_p P(p) \log_2 P(p) \quad (3.33)$$

is the entropy of a probability distribution  $P(p)$ . Hence the mutual information measures the reduction in entropy – or uncertainty – in the distribution of a variable given knowledge about another variable. Mutual information is symmetric in its arguments,  $I(p_{out}, p_{in}) = I(p_{in}, p_{out})$  and is measured in bits. An information content of  $n$  bits means that there are roughly  $2^n$  distinguishable states in our system. For instance, an input-output relation that is a step-function has 1 bit of information, as we can distinguish between a low output state and a high output state. Since the entropy or uncertainty in a variable can only decrease by measuring another variable, we have  $S[P(p_{out}|p_{in})] \leq S[P(p_{out})]$  and therefore  $I(p_{out}, p_{in}) \geq 0$ . The maximal value of  $I(p_{out}, p_{in})$  is reached when the two variables are perfectly related through a one-to-one mapping. In that case,  $P(p_{out}|p_{in}) = 0$  and  $S[P(p_{out})]$ . For our system, since we treat all  $N+1$  values for  $p_{in}$  equal, the distribution of  $p_{out}$  will be a uniform distribution,  $P(p_{out}) = 1/(N+1)$ . Therefore, the maximum value of  $I(p_{out}, p_{in})$  for our system equals  $\log_2(N+1)$ .

**Interpretation** The interpretation of mutual information can be best illustrated by the extreme examples of the ‘insulating’ phases (all ON/OFF and autonomy), in which the dynamics of the system is trivial. In the all ON and all OFF phases, all initial configurations get mapped to the same final configuration. Therefore, by knowing the final configuration we learn nothing additional about the initial configuration and vice versa. We already knew that all the cells would turn ON or OFF and knowing the initial state of the cells gives us no further information about the output. In this case,  $P(p_{out}|p_{in}) = P(p_{out})$  and the mutual information equals zero. At the other extreme, in the autonomy phase, there is a one-to-one relation between the input and output, namely  $p_{out} = p_{in}$ . In this case, knowing initial state tells us precisely what the final state is. Hence we go from a state of complete ignorance about the output, in which case there are  $N+1$  possible values for  $p$ , to having precise knowledge about the output state if we know the input state. Our ignorance about the output decreases by  $N+1$  bits by knowing the input and the mutual information between input and output states equals  $\log_2(N+1)$ . Thus, we

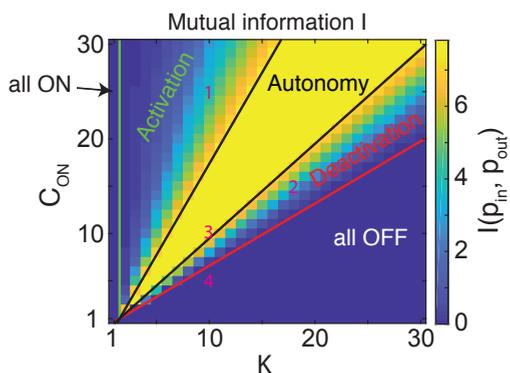
can regard mutual information as a measure for autonomy in the system. The more autonomous the system, the more each initial state corresponds to a unique final state, and hence the higher the mutual information between initial and final state.

**Relation to entropy of population** Previous work defined a quantity called “entropy of population”, which is a measure of the number of equilibrium states [Maire & Youk, 2015a]. More precisely, if  $\Omega_E$  is the number of equilibrium states, the entropy of population is defined as  $S = \log \Omega_E$ . This quantity directly measures the number of microstates of the system that are in equilibrium, and hence our ignorance of the equilibrium states if we only know the parameters of our system ( $N, a_0, K, C_{ON}$ ). As such, it is also a measure of the autonomy of the system. If the system is autonomous, there are many equilibrium states and the uncertainty about the final state is high. Conversely, if the system acts highly collectively, the final state becomes predictable and the entropy of population is low. However, as we just discussed, mutual information has a similar interpretation in terms of the degree of autonomy (vs collectiveness) of our multicellular system.

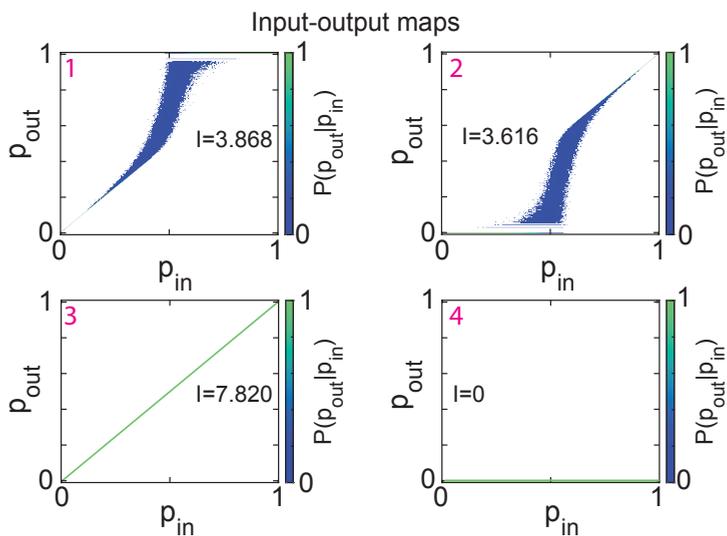
It thus becomes evident that mutual information and entropy of population are two related concepts. This can be made explicit by plotting both quantities for the various phenotypic phases of the system, controlled by  $K, C_{ON}$  (Fig. 3.11A, C). As we already derived, both quantities are minimal in the purely collective phases (all ON/OFF) and maximal in the fully autonomous phase (autonomy). In the active phases (activation/deactivation) between these extremes, both quantities are gradually modulated from one value to the other. We can take the values of both quantities we computed at each value of  $K, C_{ON}$  and directly compare them. We then find a clear positive relation between them (Fig. 3.11D) — if the entropy of population increases, the mutual information also increases and vice versa. This trend is not perfectly uniformly increasing, and this is most likely due to the fact that both quantities are computed on a macrostate level, taking only into account  $p_{in}$  and  $p_{out}$ . In this process, some information about the microstates is lost.

If both mutual information and entropy of population can be regarded as quantitative order parameters characterizing the degree of autonomy versus collectiveness in the system, then when should we prefer one over the other? The key difference is that the entropy of population weighs all equilibrium states equally, whereas the mutual information does not. Suppose that there are only two possible equilibrium states. If one equilibrium state is reached for half of the values of  $p_{in}$  and the other for half of the values, then the mutual information is 1 bit, since measuring  $p_{in}$  will give certainty over which of the two values of  $p_{out}$  is the output. However, if for each value of  $p_{in}$ , both outcomes are equally likely, then measuring  $p_{in}$  gives no information about the output

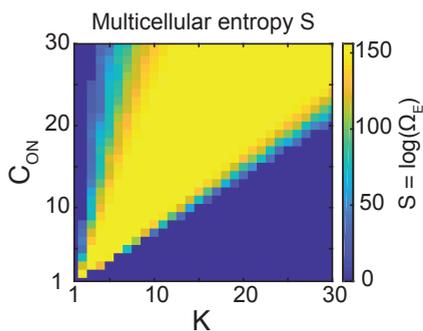
A



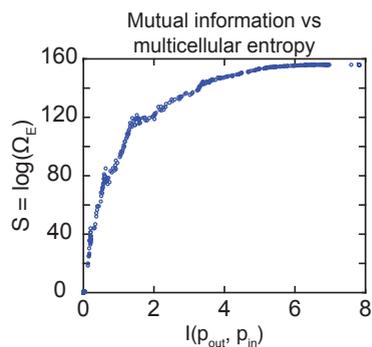
B



C



D



and the mutual information is 0 bits. The entropy of population for both cases is  $\log(2)$ . This scenario does not occur in our system, since the distribution of the output  $p_{out}$  is never independent of  $p_{in}$ , except in the case of the all ON/OFF phases (in which case both quantities yield zero). Thus, entropy of population is only a measure for the diversity of possible equilibrium states in the system, but does tell us anything about how likely these states are reached under various conditions. Mutual information does take this into account, but can give relatively low values if there are many equilibrium states which can only be reached for a small range of  $p_{in}$  values, such is often the case in the activation and deactivation regimes (see for instance panel 2 of Fig. 3.12B).

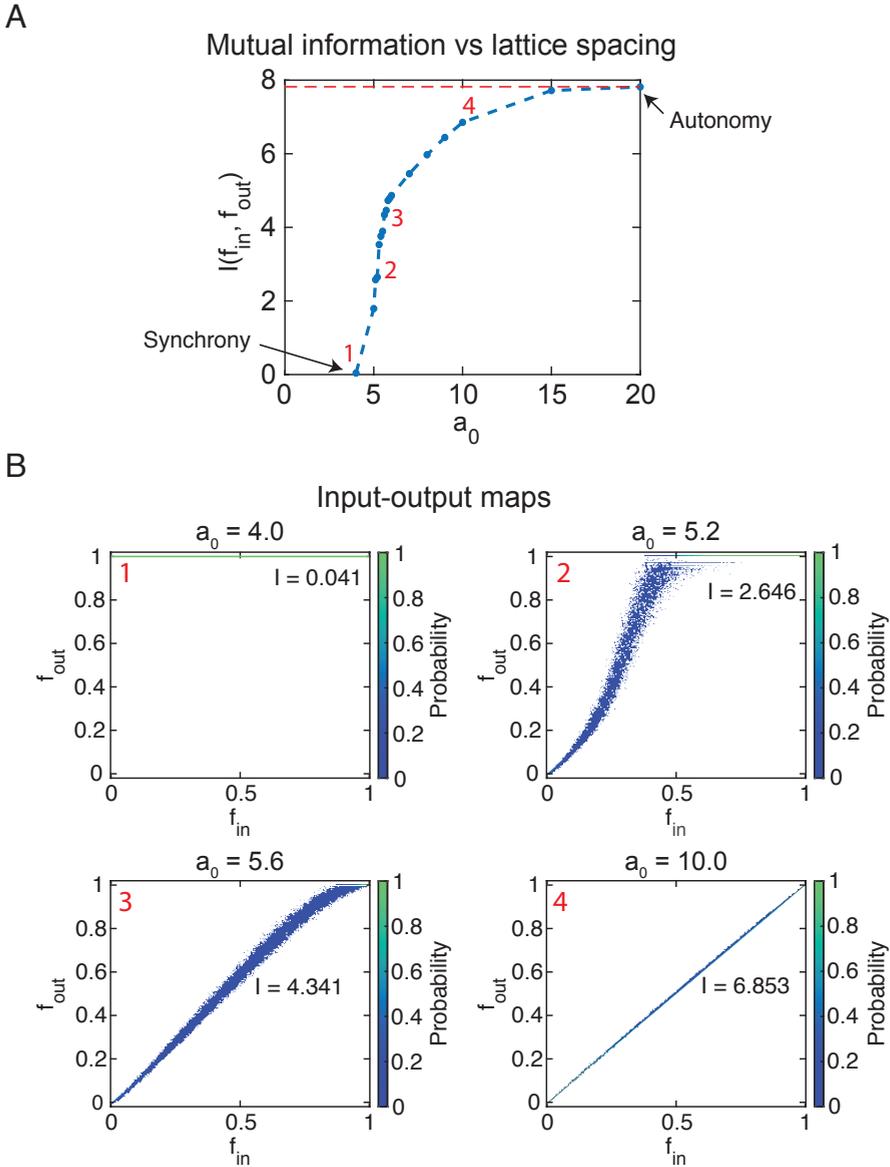
**Reversing dynamics** Why do we take a coarse-grained approach and build up an empirical probability distributions such as  $P(p_{out}|p_{in})$  which can then be analyzed through information theory and allow for Bayesian inference, in order to tell something about the initial state of the system? It seems more intuitive to simply run simulations backward in time to see how it started. This is unfortunately not possible in general, i.e. the dynamics of cellular automata is usually reversible, because configurations are typically not mapped one-to-one. Given a certain configuration  $\{X_i\}_{i=1}^N$ , there could be many past trajectories that lead to this microstate. Hence, from any configuration there might be an arbitrary number of trajectories that lead to that state (limited only by the system size). Conversely, there could be states that have no predecessor states; such states are called Garden of Eden states. Finding these Garden of Eden states is highly non-trivial, and in fact it cannot be decided whether such states exist for any type of cellular automaton with dimension higher than one [Kari, 1994].

### 3.4.2. CONTINUOUS SYSTEM

Since mutual information is a measure for the degree of autonomy in the system, it should be usable to distinguish between the collective synchronization and the autonomous regimes in the system with finite Hill coefficient. Let us consider the mutual information

---

**Figure 3.11 (preceding page): Mutual information between input and output states for a system of binary cells depends on the regions of the phenotype diagram.** (A) Mutual information between initial and final fraction of ON cells across the various regions of the phenotype diagram, with fixed parameters  $N = 225, a_0 = 1.5$  (weak interaction). The corresponding value for the maximum value of  $I$  is  $\log_2(N + 1) \approx 7.8202$ . (B) Examples of  $p_{in} - p_{out}$  maps with different values for mutual information. Mutual information is minimal in the all ON/OFF phase and maximal in the autonomy phase. The values of  $K, C_{ON}$  for each of the four labeled subfigures are indicated by the numbers on the plot in (A). (C) Entropy of population  $S = \log \Omega_E$  for the parameter range of  $K, C_{ON}$  and same values of  $N, a_0$  as in (A). (D) Relation between mutual information and entropy of population. Each point shows a point from the heat map in (A) corresponding to a set of values  $(K, C_{ON})$ . Entropy of population is computed from a simulation that estimates probabilities that the system is in equilibrium for given fraction of ON cells by trying random initial configurations.



**Figure 3.12: Mutual information as a function of the lattice spacing  $a_0$  for a system of continuous cells.** Parameters are fixed at  $N = 225, K = 8, C_{ON} = 16, n = 2, \alpha = 0$ . (A) Mutual information against  $a_0$  shows the transition from the collective phase ( $I = 0$ ) to the autonomous phase  $I = I_{max} = \log_2(N + 1)$ . The red line indicates the value of  $I_{max}$ . (B) Examples of  $p_{in} - p_{out}$  maps at different levels of noise. Results are for the activation-deactivation-phase with the same parameters as above.

$I(f_{in}, f_{out})$  between fractions of ‘ON’ cells in the binary description for  $n = 2$ , a choice that we will motivate below. In the synchronization phase, all cells acquire the same final state and  $f_{out} = 0$  or  $f_{out} = 1$ , and therefore the mutual information  $I = 0$ . In the autonomy phase, each cell remains within its basin of attraction, so the  $f_{in} = f_{out}$  and the mutual information is maximal at  $I = \log_2(N + 1)$ . In between these regimes, it takes intermediate values in these extremes. Therefore, by knowing the value of the mutual information only, we can determine whether the system is in the collective synchronization phase, the autonomy phase or the transition phase in between.

In fact, the mutual information appears to increase monotonically from zero to its maximum value as  $a_0$  is increased in this transition regime (Fig. 3.12(A)). This is reflected by the fact that the input  $f_{in}$  and output  $f_{out}$  get closer and closer together as  $a_0$  is increased (Fig. 3.12(B)). At the onset of the transition region, there is still a wide range of  $f_{in}$  values for which the system always synchronizes at  $f_{out} = 1$  ( $a_0 = 5.2$  panel in Fig. 3.12(B)). However, at the end of the transition region, close to the autonomous regime, the system is already highly autonomous, with only at most a few cells changing their state ( $a_0 = 10$  panel in Fig. 3.12(B)). Not only the mutual information, but also the  $f_{in} - f_{out}$  maps appear to change smoothly as  $a_0$  is increased. Altogether, this implies that mutual information between input and output states could be interpreted as another order parameter characterizing the transition between the collective synchronization phase and the autonomy phase.

**Binary description** There are several technical reasons why we work in the binary description of the system, with  $f_{in}$  and  $f_{out}$  instead of the mean expression levels  $p_{in}$  and  $p_{out}$ , although we obtain qualitatively similar trends with the latter choice (not shown). Firstly, we have seen that in equilibrium, the cells are clustered tightly around two equilibrium values which we identified as ‘OFF’ and ‘ON’, thus making an effective binary description feasible. In the initial state, we initiate cells at these two equilibria and apply the randomization algorithm described in Section 3.3.2. This randomization process will cause a spread in the cell states, so the cells may not start close to these fixed points. Nevertheless, by comparing the  $f_{in} - f_{out}$  maps to those obtained from initial states with  $X_k = 0$  or  $X_k = 1$  for all cells, we find no distinguishable difference between these maps (not shown) and therefore it seems that the output  $f_{out}$  is mostly determined by the value of  $f_{in}$  only. However, the same cannot be said if we work with  $p_{in} - p_{out}$ . In contrast, it appears that fixing  $p_{in}$  with different procedures leads to highly different  $p_{in} - p_{out}$  maps. This is intuitive as can be illustrated by an example. Suppose we initiate all cells in the same state  $X_k = p_{in}$  at one of the fixed points. In the autonomous phase, all cells would remain there and  $p_{out} = p_{in}$ . However, if we instead split the cells into a group

of cells with  $X_k = 0$  and another group with  $X_k = 1$ , such that on average they still have the same value of  $p_{in}$ , then we would get a very different result. The cells in the other basin of attraction would equilibrate at the other fixed point and the system would have a different value for  $p_{out}$ . Thirdly, working with continuous gene expression would in theory require working with continuous probability distributions to calculate the mutual information. In practice, however, we can only build up these distributions from discrete data, and therefore this would require us to discretize our probability distributions. However, the discretization scheme we choose has an influence on the value of the mutual information we find. Given a discretization with  $N_b$  bins, the mutual information is bound by the maximal value  $\log_2(N_b)$ . of course, a natural choice would be to take  $N_b = N + 1$  bins, with edges at  $pN = [0, \frac{1}{2}, \frac{3}{2}, \dots, N - \frac{1}{2}, N]$ . This roughly corresponds to a 'resolution limit' of or ability to distinguish a single cell that is either completely ON or completely OFF.

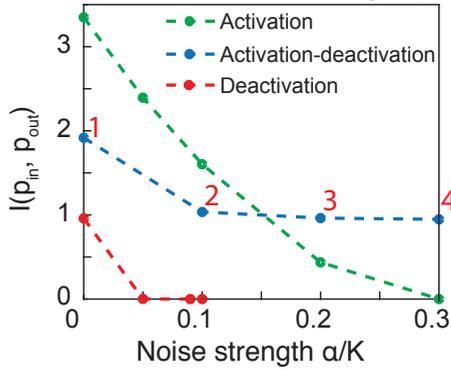
### 3.4.3. APPLICATION: ENSEMBLE-LEVEL RESPONSES TO STOCHASTIC NOISE

In this last part, we consider an application of the above framework to study the effect of stochastic noise on ensemble-level features of the system, in particular it's  $p_{in} - p_{out}$  maps. In Chapter 4, we will see that noise is capable of 'liberating' metastable states and drive them further down the pseudo-energy landscape. We also propose that this could lead to trajectories with higher levels of spatial organization, because of the inherent tendency of trajectories to acquire higher levels of the spatial index  $I$  as they evolve in time. However, these findings have been discussed only at the level of individual trajectories, and an ensemble-level discussion remains lacking. Here, we will consider sets of simulations at different values of initial  $p_{in}$  or  $f_{in}$  and look at how noise affects their output state.

**Binary system** We first considered the binary system and study it's ensemble-level responses by constructing  $p_{in} - p_{out}$  maps through performing many simulations. The addition of noise is most interesting in the active phases of the system (activation, deactivation and activation-deactivation), as we expect to see higher levels of (de)activation as noise level is increased. Indeed, if the system is in (de)activation phase, adding moderate noise will only tend to drive the system towards more (de)activation and not cause any cells to deactivate (activate). However, one effect that follows from this is that the set of metastable states, where the system does not fully (de)activate and tends to end up in configurations of high spatial organization, tends to get reduced as noise is increased. Both the range of  $p_{in}$  values for which metastable states are possible and the extend of possible  $p_{out}$  values for these metastable states are decreased. This process is illustrated for the activation-deactivation phase in Fig. 3.13(B). Whereas for low noise the system

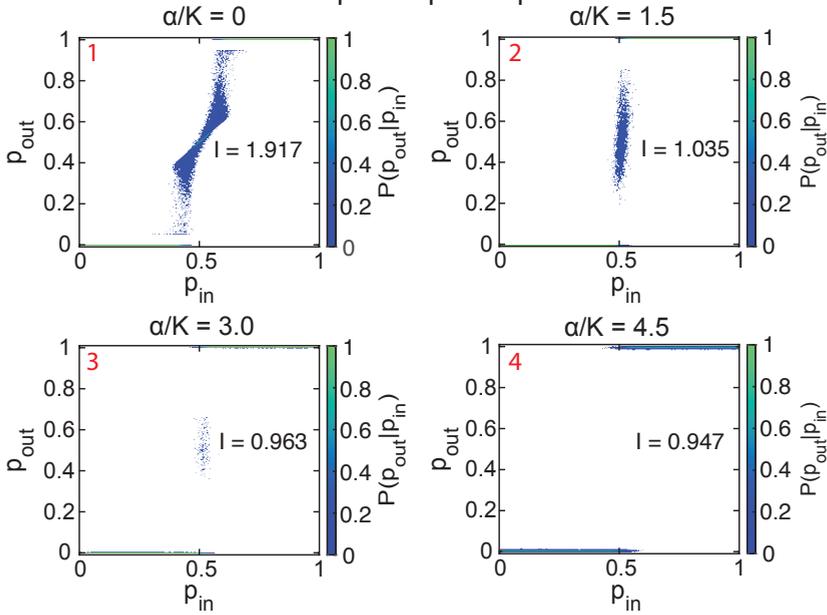
A

Mutual information vs noise strength (binary cells)



B

Input-output maps



**Figure 3.13: Influence of noise on mutual information between input and output states, for a system of binary cells.** We fixed  $N = 225$ . (A) Mutual information against noise for three different behavioral phases of the model. Red corresponds to the deactivation phase ( $a_0 = 1.5, K = 8, C_{ON} = 15$ ), blue to the activation-deactivation phase ( $a_0 = 0.5, K = 15, C_{ON} = 8$ ) and green to the activation phase ( $a_0 = 1.5, K = 20, C_{ON} = 15$ ). Note that for activation and deactivation,  $I \rightarrow 0$  as there is only one final state (all ON/OFF), whereas for activation-deactivation  $I \rightarrow 1$  bit as there are two stable states which could only potentially mix at extremely high values of noise. (B) Examples of  $p_{in} - p_{out}$  maps in the activation-deactivation phase at different values of the noise strength.

can still end up in some metastable states if  $p_{in} \approx 0.5$ , at higher levels the response becomes binary: either all cells turn OFF (if  $p_{in} < 0.5$ ) or all cells turn ON (if  $p_{in} > 0.5$ ). Correspondingly, the mutual information between  $p_{in}$  and  $p_{out}$  decreases uniformly for all three studied phases as noise strength is increased (Fig. 3.13(A)). In the activation and deactivation phases,  $I(p_{in}, p_{out})$  goes to zero, as all trajectories get driven to the single global minimum. In the activation-deactivation phase, noise drives the system to either  $p_{out} = 0$  or  $p_{out} = 1$  depending on the initial state. Knowing  $p_{in}$  then allows one to distinguish between these two cases, so  $I(p_{in}, p_{out}) \rightarrow 1$ .

**Continuous system** In the continuous system, we observed a similar phenomenon of noise leading to more uniform end states and thereby decreasing mutual information. We considered the system with parameters in the transition regime, close to the autonomous phase, such that without noise a variety of final states are possible. With the addition of noise, the system is driven towards the higher fixed point, as is evident from the  $f_{in} - f_{out}$  maps in Fig. 3.14A. Why does it seem that noise biases the cells towards the higher fixed point (since  $f_{out} > f_{in}$ )? This has to do with the fact that location of the unstable fixed point, however, is typically not precisely in the middle between the two stable fixed points (found by solving Eq. 3.16). In particular, for our set of parameters the basin of attraction of the lower fixed point is much smaller than that of the higher fixed point. As a result, noise from one of the fixed points is more likely to push the system over to the other fixed point than vice versa. Therefore, at relatively high levels of noise, all trajectories go to a  $f_{out} = 1$  final state and mutual information goes to zero ( $\alpha = 10^{-0.3}$  panel in Fig. 3.14B). As noise is further increased, however, some transitions back to the low fixed point are possible and the system tends to stabilize in a dynamic steady state with  $f_{out}$  values close to but less than 1 ( $\alpha = 10^{0.5}$  panel in Fig. 3.14B). This corresponds to the slight increase in  $I(f_{in}, f_{out})$  in Fig. 3.14A. Given these results together, we deduce that there are three regimes upon increasing noise in the system. Note that this is a phenomenological description and that the boundaries between the regimes are not strictly defined.

1. In the first regime, noise has limited effect on the expression level of individual cells and hardly causes any cells to switch state. Noise has the effect to cause fluctuations in the states of individual cells, but none of these fluctuations lead to large effects that spread across the entire system.
2. In the second regime, noise can induce only one-sided transitions from the state closest to the unstable fixed point. In our example above, cells can easily turn from OFF to ON, but not the other way around. This causes the system to become completely homogeneous over time, as all tells are pushed towards the fixed point

with the larger basin of attraction.

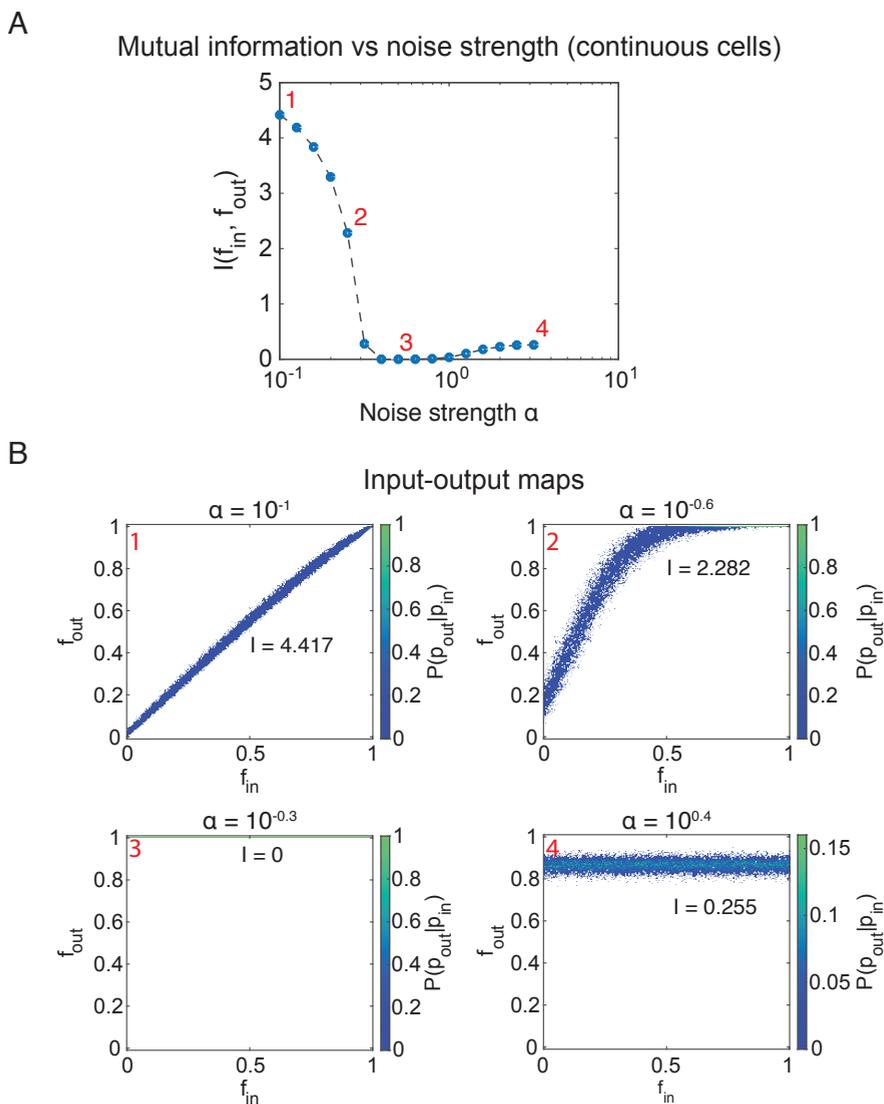
3. In the third regime, noise can induce both transitions from the low to the high expression level fixed point, and vice versa. Because of the high level of noise, the final steady is a dynamical steady state where the gene expression levels still fluctuate over time, but on average tends to be within a given range of  $f_{out}$  values.

Altogether, these results show that adding noise to system with continuous response functions leads to more uniformity in the final pattern that form. This is in contrast to the discrete model, for which moderate noise can lead to more ordered final states, as will be discussed in Chapter 4 (for one molecule) and Chapter 5 (for two molecules).

#### 3.4.4. DISCUSSION

In this section, we characterized input-output relations between the initial and final states of the system through the information-theoretic concept of mutual information. We have shown how mutual information relates to the degree of autonomy in the system. Its value is low when different initial configurations give rise to the same final configuration, and high when they lead to unique final states. As such, mutual information is related to the ‘entropy of population’ defined in earlier work [Maire & Youk, 2015a], as we have also shown numerically. However, the main advantage of using mutual information over ‘entropy of population’ is that it takes into account the fact that not all equilibrium states occur with equal probability.

From an information perspective, mutual information expresses the number of bits of information that can be learned about one variable by measuring the other remains true. As such, it is strongly related to the notion of correlation as measured by well-established correlation coefficients. However, whereas correlation typically refers to linear dependencies between variables, mutual information represents a fully nonparametric method of assessing potentially nonlinear relationships between variables. As such, the results we obtained cannot have been obtained by substituting mutual information by correlation coefficient.



**Figure 3.14: Mutual information as a function of the noise strength  $\alpha$  for a system of continuous cells.** Parameters are fixed at  $N = 225, K = 8, C_{ON} = 16, n = 2, a_0 = 6$ . Note that without noise, this system would lie in the autonomous phase. (A) Mutual information against noise strength shows three different regimes. First, mutual information decreases with noise as the system is driven towards homogeneous final states. There is a plateau at  $I = 0$  for intermediate noise levels, corresponding to the situation where all cells turn ON over time due to noise. Finally, at high noise levels the ON cells are no longer stable and stochastically transition to OFF states. (B) Examples of  $p_{in} - p_{out}$  maps at different values of the noise strength show these regimes explicitly.

# 4

## STATISTICAL DYNAMICS OF SPATIAL-ORDER FORMATION BY COMMUNICATING CELLS

*Communicating cells can coordinate their gene expressions to form spatial patterns, generating order from disorder. Ubiquitous “secrete-and-sense cells” secrete and sense the same molecule to do so. Here we present a modeling framework — based on cellular automata and mimicking approaches of statistical mechanics — for understanding how secrete-and-sense cells with bistable gene expression, from disordered beginnings, can become spatially ordered by communicating through rapidly diffusing molecules. Classifying lattices of cells by two “macrostate” variables — “spatial index”, measuring degree of order, and average gene expression level — reveals a conceptual picture: a group of cells behaves as a single particle, in an abstract space, that rolls down on an adhesive “pseudo-energy landscape” whose shape is determined by cell-cell communication and an intracellular gene-regulatory circuit. Particles rolling down the landscape represent cells becoming more spatially ordered. We show how to extend this framework to more complex forms of cellular communication.*

---

An adapted version of this chapter has been published as E.P. Olimpio\*, Y. Dang\* and H. Youk, “Statistical dynamics of spatial-order formation by communicating cells”, *iScience* **2**, 27-40 (2018)  
Some results from this chapter have also appeared in the master thesis of E.P. Olimpio [Olimpio, 2016].

## 4.1. INTRODUCTION

Cells can communicate by secreting signaling molecules and this often underlies their collective behaviors. A striking example is that of initially uncoordinated cells, through cell-cell communication, coordinating their gene expressions to generate spatial patterns or structures [Gregor et al., 2010; Sawai et al., 2005; Danino et al., 2010; Liu et al., 2011]. Many cells partly or completely control such “disorder-to-order” dynamics by simultaneously secreting and sensing the same signaling molecule [Doğaner et al., 2016; Youk & Lim, 2014a]. These “secrete-and-sense cells” appear across diverse organisms and include quorum-sensing social amoeba, *Dictyostelium discoideum*, that form fruiting bodies [Gregor et al., 2010; Sawai et al., 2005; Sgro et al., 2015] and autocrine-signaling T-cells [Antebi et al., 2013; Sporn & Todaro, 1980; Youk & Lim, 2014b]. Based on mounting evidence from studies of various organisms [Gregor et al., 2010; Danino, et al., 2010; Youk & Lim, 2014a; Antebi et al., 2013; Mehta et al., 2009; Kamino et al., 2017; Hart et al., 2014; De Monte et al., 2007; Umeda and Inouye, 2004; You et al., 2004; Pai et al., 2012; Coppey et al., 2007; Shvartsman et al., 2001], researchers now suspect that secrete-and-sense cells, many of which are governed by the same type of genetic-circuit [Doğaner et al., 2016], are highly suited for spatially coordinating their gene expressions. But if true, exactly why this is so, whether there are common design principles shared by the different organisms, what the dynamics underlying their disorder-to-order transition is, and how to even quantify their spatial order, remain open questions. In this chapter, we address these questions in the context of initially disordered fields of secrete-and-sense cells that self-organize into spatially ordered fields without any pre-existing morphogens. Specifically, we develop a theoretical framework that takes a simple and ubiquitous class of secrete-and-sense cells, sensibly defines and quantifies the notion of the cells’ spatial order, and then elucidates how the spatial order evolves over time. We focus here on analytically describing how spatial correlations among cells’ gene-expression levels dynamically emerge rather than on describing the shapes, sizes, and formations of specific spatial patterns (e.g., stripes). To study how these cells generate specific patterns, one often uses exhaustive numerical simulations that are adapted to particular settings [Cotterell & Sharpe, 2010; Cotterell et al., 2015; Chen et al., 2015]. While such simulations provide insights into the dynamics of spatial-order formation, a different modeling framework may provide complementary insights that are difficult to extract from the often-large numbers of parameters involved in numerical simulations.

Our main idea is that describing hundreds to thousands of secrete-and-sense cells forming a particular spatial configuration is infeasible without exhaustive numerical simulations but that it is possible to analytically describe how an ensemble of “similar” spatial configurations evolves over time without knowing the state of every single cell. As we

will show, we do this by defining quantities that are similar to those found in statistical physics but have meanings and properties that are very different and are adapted for describing cells. Specifically, we will define a “spatial index” — a number whose magnitude is between zero (complete disorder) and one (complete order). Inspired by approaches of statistical mechanics, we will group all lattices of cells that have the same spatial order parameter and average gene-expression level into an ensemble that we will call a “macrostate”. Surprisingly, we find that this macrostate moves like a particle that drifts-and-diffuses in an abstract, two-dimensional space that we will call a “phase space” — since it describes all possible spatial configurations of the lattice — and whose coordinates denote the cells’ spatial order and average gene-expression level. We find that the particle, representing an entire cellular lattice, moves in the phase space by rolling down on a “pseudo-energy landscape”, which is a visual landscape that is shaped by the communication among the cells and the intracellular gene-regulatory circuit that controls how the cells secrete and sense the molecule. We will show that the shape of this landscape is quantitatively defined by a function that we will call a “pseudo-energy” and show that while it mathematically resembles the Hamiltonian of the Ising model, it has different properties. We will show that the gradient of the pseudo-energy and a “trapping probability”, which quantifies an adhesiveness of the pseudo-energy landscape, together determine the particle’s trajectories in the phase space — the particle rolls down along the negative of the gradient of the pseudo-energy and at locations where the landscape is highly adhesive, it halts. Crucially, we will show that these trapping locations on the pseudo-energy landscape — the locations where the particle halts — correspond to highly ordered spatial configurations such as islands of cells that have the same gene-expression level. A moderate amount of noise can induce the particle to roll down further on the pseudo-energy landscape and this corresponds to the cells forming patterns with even higher spatial organizations. We thus provide an intuitive and visual picture, based on experimentally attainable quantities, that is both practical and conceptual for elucidating how a simple class of secrete-and-sense cells spatially coordinate their gene expressions.

## 4.2. RESULTS

### 4.2.1. CELLULAR AUTOMATON SIMULATES SECRETE-AND-SENSE CELLS THAT

SLOWLY RESPOND TO A RAPIDLY DIFFUSING SIGNALING MOLECULE.

We used a cellular automaton [Ermentrout & Edelstein-Keshet, 1993] to simulate secrete-and-sense cells. We will compare the results of the cellular automaton with our theory’s predictions. We considered a two-dimensional, triangular lattice of  $N$  spherical, immobile secrete-and-sense cells of radius  $R$  and a lattice spacing  $a_0$ . As a proof-of-

principle, we considered “simple” secrete-and-sense cells which we define to be cells that (1) very slowly respond to their fast diffusing signal, and (2) whose gene-expression level, which is determined by the extracellular concentration of the signal and signal-secretion rate, exhibit switch-like (digital) bistability (see Chapter 3.1). These two features were motivated by experimentally characterized secrete-and-sense cells. Examples include yeasts that secrete-and-sense a mating pheromone in a nearly digital manner (diffusion timescale  $\sim 1$  second; response timescale  $\sim 30$  minutes) [Youk & Lim, 2014a; Rappaport & Barkai, 2012] and mouse hair follicles, which are secrete-and-sense organs that act as digital secrete-and-sense cells on a triangular lattice (diffusion timescale  $\sim 12$  hours; response timescale  $\sim 1.5$  days) [Chen et al., 2015; Maire & Youk, 2015b] (see also Table S4.1). Each cell’s gene expression is either “ON” (when its signal-secretion rate is at a maximum) or “OFF” (when its signal-secretion rate is at a minimum, basal level). Each cell senses a steady-state signal-concentration  $c$  on itself. If  $c$  is higher (lower) than a threshold concentration  $K$ , which we call an “activation threshold”, then the cell is ON (OFF). We set  $C_{OFF} = 1$  so that we express all concentrations as multiples of  $C_{OFF}$ . Our cellular automaton computes the concentration on every cell, then synchronously updates each cell’s state, and then repeats this process until the cellular lattice reaches a steady-state configuration in which no cell’s state requires an update. By running the cellular automaton on randomly distributed ON- and OFF-cells, we observed that initially disordered lattices could indeed evolve into spatially ordered steady-state configurations such as islands of ON-cells (Figure 4.1A).

#### 4.2.2. SECRETE-AND-SENSE CELLS CAN BE CLASSIFIED INTO DISTINCT BEHAVIORAL PHASES.

To reveal how the disorder-to-order dynamics arises, we will analyze the cellular automaton in each of the cells’ “behavioral phases” that we described in a previous work (Figure 4.1B) [Maire & Youk, 2015a]. As the previous work showed, the behavioral phases represent how one cell turns on/off another cell. They arise from self-communication (i.e., a cell captures its own signal) competing with neighbor-communication (i.e., a cell captures the other cells’ signal). The communication between two cells, cell- $i$  and cell- $j$ , is quantified by an “interaction term” for that pair,  $f(r_{ij}) \equiv e^{(R-r_{ij})/r_{ij}} \sinh(R)$  (where  $r_{ij}$  is the distance between the centers of cell- $i$  and cell- $j$  and  $R$  is both cells’ radius). This term is directly proportional to the concentration of the signaling molecule on cell- $i$  that is due to cell- $j$ , and vice-versa. We then quantify the competition between the self- and neighbor-communication among the  $N$  cells with the “interaction strength”,  $f_N(a_0) \equiv \sum_{i,j} e^{(R-r_{ij})/r_{ij}} \sinh(R)$ , which is the sum of the interaction terms of all cell-pairs. It is a function only of the cells’ radius  $R$  and the lattice spacing  $a_0$ . The latter is

because all distances between the cells are determined by specifying the lattice spacing. The interaction strength  $f_N(a_0)$  measures how much each cell captures the signals from all the other cells (see Eq. 3.5) [Maire & Youk, 2015a]. For a given interaction strength, the activation threshold  $K$  and the  $C_{ON}$  determine the cells' behavioral phase. The values of  $K$ ,  $C_{ON}$ , and  $f_N$  are held fixed and thus the cells' behavioral phase also remains unchanged over time. We categorize a behavioral phase as either an “insulating phase” — in which no cell can turn on/off the other cells due to dominant self-communication — or a “conducting phase” — in which cells can turn on/off the others due to dominant neighbor-communication (Figure 4.1B). Regardless of the interaction strength, cells can operate in two conducting phases: (1) “activate phase” — in which neighboring ON-cells can turn on an OFF-cell, and (2) “deactivate phase” — in which neighboring OFF-cells can turn off an ON-cell. Additionally, when the interaction is weak (i.e.,  $f_N(a_0) < 1$ ), cells can operate in an “autonomy phase”, which is an insulating phase whereby a cell can stay ON/OFF regardless of the other cells' states. On the other hand, when the interaction is strong (i.e.,  $f_N(a_0) > 1$ ), cells can operate in an “activate-deactivate phase”, which is a conducting phase whereby the cells can both activate and deactivate the others depending on their relative locations.

#### 4.2.3. GROUPING MULTIPLE SPATIAL CONFIGURATIONS INTO ONE MACROSTATE BASED ON THEIR COMMON SPATIAL INDEX $I$ AND FRACTION $p$ OF CELLS THAT ARE ON.

We now present our framework's central ingredient. Let us define two “macrostate” variables: (1) the fraction  $p$  of cells that are ON (equivalent to the average gene-expression level) and (2) a “spatial index”  $I$  that we define as

$$I = \frac{N}{\sum_i \sum_{j \neq i} f(r_{ij})} \frac{\sum_i \sum_{j \neq i} f(r_{ij}) (X_i - \langle X \rangle) (X_j - \langle X \rangle)}{\sum_i (X_i - \langle X \rangle)^2} \quad (4.1)$$

where  $X_i$  is +1(−1) for an ON (OFF)-cell and  $\langle X \rangle = \sum_i X_i / N$  is the average over all the cells. The spatial index  $I$ , in fact, belongs to a widely-used class of statistical metrics called Moran's I [Moran, 1950]. Moran's I is frequently used for spatial analysis in diverse fields, including geographical analysis [Getis & Ord, 1992], ecology [Legendre, 1993] and econometrics [Anselin, 2008]. Our spatial index  $I$  measures a spatial autocorrelation among the cells by weighing each cell-pair by that pair's interaction term  $f(r_{ij})$  [Maire & Youk, 2015a]. Thus, roughly speaking, the spatial index measures the average correlation between the states of any two cells by assigning a higher weight to those cell-pairs that communicate more with each other (and hence are closer to each other in space). By construction,  $-1 \leq I \leq 1$  and  $0 \leq p \leq 1$ . When  $I = 0$ , ON- and OFF-cells are randomly

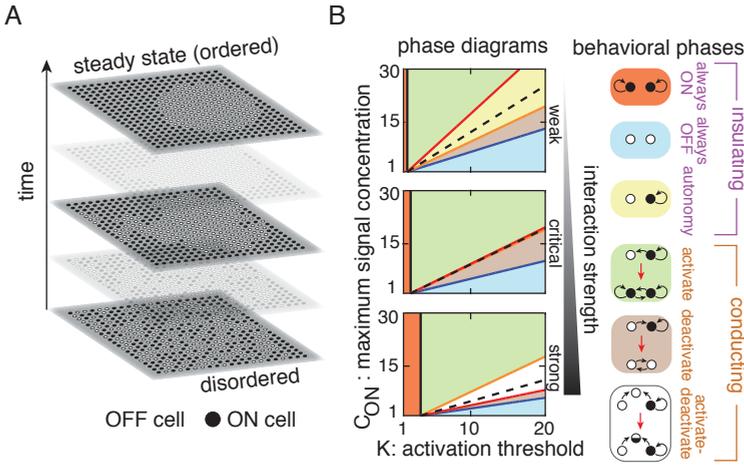
distributed across the lattice, yielding maximally disordered lattices (Figure 4.2A – top row). When  $|I|$  is large, the cells are more spatially ordered and the lattice consists of large contiguous clusters of ON/OFF-cells (Figure 4.2A – bottom row). For  $I > 0$ , cells of the same ON/OFF-state tend to cluster together, whereas for  $I < 0$ , cells of the same ON/OFF-state tend to avoid each other. As we will see below, we can focus on lattices with a positive spatial index for our purpose. For positive values of  $I$ , a key feature that the value of the spatial index tells us is whether the lattice consists of one large, contiguous island of ON/OFF-cells (when  $I$  is close to one; Figure 4.2A – bottom row) or of many fragmented small islands of ON/OFF-cells (when  $I$  is close to zero; Figure 4.2A – top row).

## 4

Our central idea is to group cellular lattices that have the same  $(p, I)$  into a single ensemble (examples in Figure 4.2A). We then view this ensemble as a particle that moves in an abstract space whose position at time  $t$  is  $(p(t), I(t))$ . We call this abstract space a “phase space” because each point  $(p, I)$  in this space represents an ensemble of all possible spatial configurations that have the same value of  $p$  and the same value of  $I$ . The number of microstates that share the same macrostate can then be computed using methods based on those used to calculate the density of states in statistical physics (Section S4.5.3). The procedure of grouping spatial configurations based on their  $(p, I)$  is akin to a situation in physics in which many microstates (e.g., the positions and momenta of all particles) are grouped into a single macrostate (e.g., pressure or temperature). Thus, we will call each lattice configuration a “microstate” and the ensemble of these microstates represented by a given  $(p, I)$  to be a “macrostate” (Figure 4.2A).

#### 4.2.4. CELLULAR LATTICE IS REPRESENTED BY A PARTICLE WHOSE POSITION $(p, I)$ AND TRAJECTORY DEPENDS ON THE BEHAVIORAL PHASE.

By randomly choosing thousands of microstates that all belong to the same disordered macrostate ( $p = p_{\text{initial}}, I \approx 0$ ) and then running the cellular automaton on each of these microstates, we observed how the lattices evolved out of disorder. Specifically, we obtained a distribution of their trajectories, and thus also a distribution of their final positions ( $p = p_{\text{final}}, I = I_{\text{final}}$ ), for every value of  $p_{\text{initial}}$  in each behavioral phase (Figure 4.2B). The fact that we obtained, for a fixed value of  $p_{\text{initial}}$ , a distribution of values for  $p_{\text{final}}$  (Figure 4.2B – top row) and a distribution of trajectories (Figure 4.2B – bottom row) instead of a single trajectory, indicate that the particle moves stochastically in the  $p$ - $I$  space. This stochasticity arises from the cellular automaton operating on individual cell's state  $X_k$ , a microstate variable, at each time step rather than operating on the macrostate variables,  $p$  and  $I$ . And since, at the macrostate-level, we are ignorant of the exact microstate that the cellular automaton is operating on, the macrostate-level description of



**Figure 4.1: Secreted-and-sense cells can be classified into distinct behavioral phases.** (A) Snapshots of cellular automaton at different time points, in which an initially disordered cellular lattice becomes more ordered over time. White circle is an OFF-cell and a black circle is an ON-cell. (B) (Left column): Phase diagrams for a weak interaction (top panel;  $f_N(a_0) < 1$ ), a critical interaction (middle panel;  $f_N(a_0) = 1$ ), and a strong interaction (bottom panel;  $f_N(a_0) > 1$ ), where the interaction strength is defined as  $f_N(a_0) = \sum_{i,j} \frac{e^{R-r_{ij}}}{r_{ij}} \sinh(R)$ . (Right column): Different colors denote distinct behavioral phases. See also Table S4.1.

the particle's motion, once we deduce it, would have to be a stochastic description. We found several promising signs that an analytical, macrostate-level description is possible. Firstly, we observed that particles that started at the same position ( $p_{\text{initial}}, 0$ ), for the most part, remained close to each other in subsequent times, leading to tightly bundled trajectories in the  $p$ - $I$  space despite the stochasticity (Figure 4.2B – bottom row). Furthermore, we observed other features that were shared by all the trajectories for each behavioral phase. Specifically, in the activate phase, we observed that if the  $p_{\text{initial}}$  was above a certain threshold value (red vertical line in Figure 4.2B – top left panel), then almost all cells were turned on whereas if it was below the threshold value, then the activation was minimal due to the cellular automaton not starting with enough ON-cells. In the deactivate phase, we observed that if the  $p_{\text{initial}}$  was below a certain threshold value (red vertical line in Figure 4.2B – top middle panel), then almost all cells turned off whereas if it was above the threshold, then the deactivation of ON-cells was minimal due to the cellular automaton not starting with enough OFF-cells. Finally, in the activate-deactivate phase, we observed a threshold value for activation (green vertical line in Figure 4.2B – top right panel) and a threshold value for deactivation (red vertical line in Figure 4.2B – top right panel). Between these two thresholds, a particle stops with a value of  $p$  that is either only slightly higher (activation) or slightly lower (deactivation)

than the value that it started with (giving rise to a slanted “bowtie” shape in Figure 4.2B – top right panel).

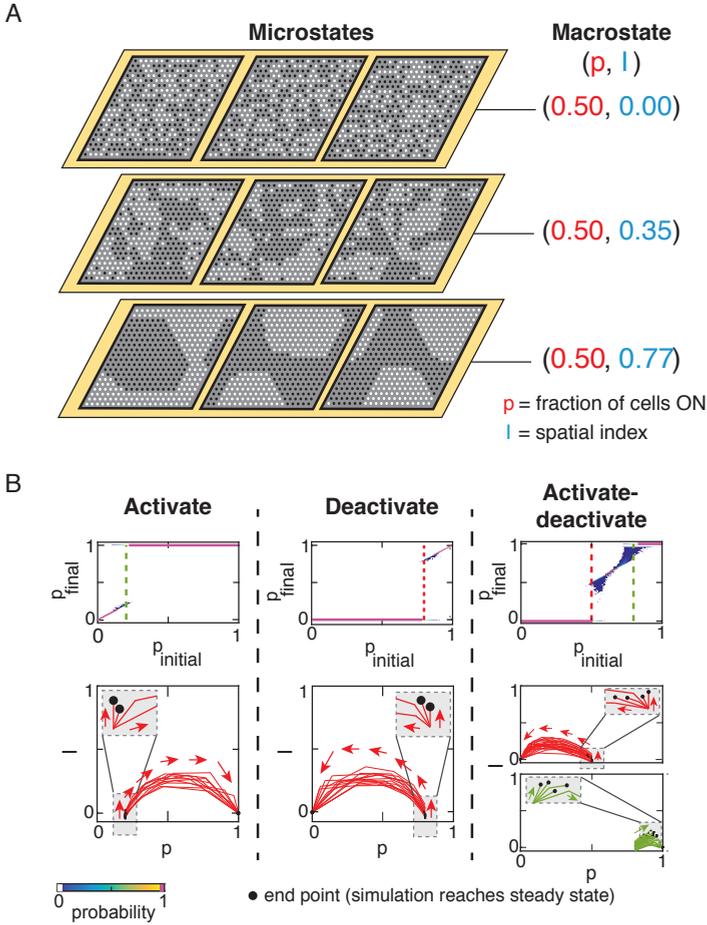
We also observed common features in the shapes of the trajectories themselves in the  $p$ - $I$  space. Specifically, we observed that in every trajectory, the  $I$  initially increased before plateauing at some value while the  $p$  either monotonically increased or decreased over time (Figure 4.2B – bottom row). Then, one of two events occurred in all trajectories: either (A) the particle stopped, and thus the cellular automaton terminated, with the final value of  $p$  (i.e.,  $p_{\text{final}}$ ) between zero and one (see black dots that mark the trajectories’ endpoints in Figure 4.2B – bottom row) or (B) the particle kept increasing or decreasing its  $p$  until it reached and stopped at either  $p = 1$  (all cells ON) or  $p = 0$  (all cells OFF) and as it did so, its spatial index abruptly dropped to zero (e.g., most of the red trajectories in Figure 4.2B). Observation (A) corresponds to a situation in which the cells form an ordered spatial configuration that, being a steady state of the cellular automaton, remains unchanged indefinitely. This situation arose most notably but not exclusively in the activate-deactivate phase. Observation (B) corresponds to a situation in which all cells either turn on or off.

To explain observation (B), we first rewrite Equation 4.1 as (Supplemental Information section S4.5.1)

$$I(p) = \frac{(\Theta - (2p - 1)^2 f_N(a_0))}{4p(1 - p)f_N(a_0)}, \quad (4.2)$$

where  $\Theta = \frac{1}{N} \sum_{i,j \neq i} f(r_{ij}) X_i X_j$ . Note that the  $p$  and the spatial index  $I$  depend on each other. And since Equation 4.2 enables us to deduce  $\Theta$  if we know  $I$  and vice versa, we have the option of considering  $(p, \Theta)$  to be a macrostate instead of  $(p, I)$ . The main disadvantage of this is that the  $\Theta$ , unlike the spatial index, it is not normalized. This makes it difficult to compare the values of  $\Theta$  for lattices with different values of  $p$ . Thus, we will work with  $(p, I)$  instead of  $(p, \Theta)$ . From a mean-field approximation, in which we calculate the average amount of signal sensed by each cell (Supplemental Information section S4.5.2), we deduced that the particle’s spatial index has an upper bound for each value of  $p$ . We denote this  $p$ -dependent maximal value of  $I$  by a function  $I_{\text{max}}(p)$  (dashed black curves in Figure 4.3). The function  $I_{\text{max}}(p)$  sharply drops to zero as  $p$  nears zero or one. Accordingly, as the particle’s  $p$  nears zero or one, its spatial index should sharply decrease to zero in accordance with observation (B) (Figure 4.3B-D). This makes sense because the spatial index is a measure of whether or not the lattice consists of a large, contiguous island of ON/OFF-cells. As the spatial index approaches zero, the lattice becomes populated with more fragments of smaller islands of ON/OFF-cells. When the  $p$  is near zero (one), as is the case when only one cell is ON (OFF), then no clusters of ON-cells (OFF-cells) are possible since there is only one ON-cell (OFF-cell). Due to this reason and from a rigorous calculation of how the  $I$  changes as the  $p$  approaches zero or

one (Supplemental Information section S4.5.1), we find that the spatial index is indeed zero when the  $p$  is either zero or one. To fully explain the particle trajectories along with observations (A) and (B), we next sought an equation of motion for the particles.



**Figure 4.2:** Spatial configurations of secrete-and-sense cells (microstates) can be grouped into macrostates. (A) Examples of microstates that have the same fraction of cells being ON (denoted  $p$ ) and spatial index  $I$  grouped into a single macrostate, denoted by  $(p, I)$ . For each macrostate  $(p, I)$ , three microstates are shown as examples. (B) (Top row): Probability density maps showing the particle's final value of  $p$  (denoted  $p_{\text{final}}$ ) for each initial value of  $p$  (denoted  $p_{\text{initial}}$ ) in the activate phase, deactivate phase, and activate-deactivate phase. Color code for the probability density is shown in the color bar at the bottom. The green dashed lines in the activate phase (left panel) and in the activate-deactivate phase (right panel) approximate the lowest value of  $p_{\text{initial}}$  that is required to turn on every cell (i.e., reach  $p_{\text{final}} = 1$ ). The red dashed lines in the deactivate phase (middle panel) and in the activate-deactivate phase approximates the highest value of  $p_{\text{initial}}$  required to turn off every cell (i.e., reach  $p_{\text{final}} = 0$ ). (Bottom row): Trajectories (red and green curves) in  $p$ - $I$  space (called "phase space") in the activate phase (left panel), deactivate phase (middle panel), and activate-deactivate phase (right panel). Grey insets show zoomed-in views of some trajectories. Black dots denote the trajectories' endpoints.

#### 4.2.5. CELLULAR LATTICE ACTS AS A PARTICLE THAT ROLLS DOWN ON AND ADHERES TO A PSEUDO-ENERGY LANDSCAPE

We conjectured that if a cellular lattice indeed moves like a particle, then there may be a “landscape” on which the particle rolls down. To explore this idea, we consider a function  $h$  that we call a “pseudo-energy” and define it as  $h \equiv -\sum_i X_i(Y_i - K)$ , where  $Y_i$  is the signal concentration on cell- $i$ . In fact, we can rewrite  $h$  entirely in terms of the macrostate variables,  $p$  and  $I$  (Supplemental Information section S4.5.4). Plotting  $h(p, I)$  yields a three-dimensional landscape that we call a “pseudo-energy landscape” (Figure 4.3A). Its shape depends on the cells’ behavioral phase (Figures 4.3B–D). Importantly, by plotting the trajectories on top of their respective landscapes, we observed that every particle’s pseudo-energy (i.e., value of  $h$ ) monotonically decreased over time until the particle stopped. We could also rigorously prove this (Supplemental Information section S4.5.4). The fact that the pseudo-energy is a decreasing function of the spatial index explains why trajectories in general tend towards increasing values of the spatial index (Figures 4.3B–D).

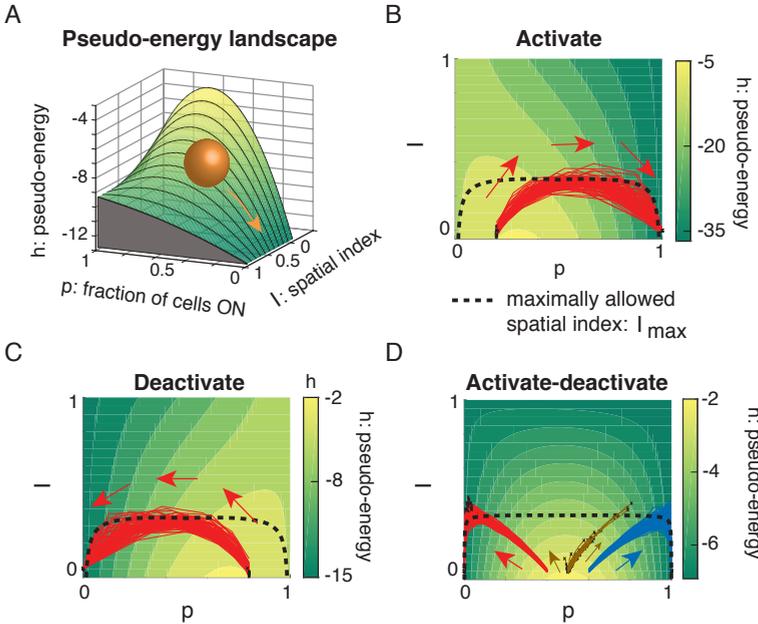
To see, at the microstate-level, why the cells’ states become more spatially correlated over time, we rewrite the  $h$  as

$$h = -\alpha \sum_{i,j \neq i} f(r_{ij})X_iX_j - B \sum_i X_i - N\alpha \quad (4.3)$$

where  $\alpha \equiv (C_{ON} - 1)/(2N)$  and  $B$  is a “signal field” defined as  $\alpha(1 + f_N(a_0)) - K/N$ . Equation 4.3 is strikingly similar to the Hamiltonians of the Hopfield network [Hopfield, 1982] and magnetic spins with long-range interactions [Kirkpatrick & Sherrington, 1975; Tchernyshyov & Chern, 2011]. Note that since  $\alpha f(r_{ij}) > 0$  and the particle’s pseudo-energy keeps decreasing over time before the particle stops, the cells must “align” their states with each other rather than “anti-align” (i.e., the pseudo-energy favors the pairing of two ON-cells rather than pairing of an ON-cell with an OFF-cell). In magnetic spin systems, this would be analogous to a ferromagnetic interaction. As in physical systems, we can view the signal field  $B$  as a macroscopic knob that we can tune to change the shape of the pseudo-energy landscape for a given cellular lattice. From the phase diagrams (Figure 4.1B), we can deduce that  $B > 0$  in the activate phase, that  $B < 0$  in the deactivate phase, and that  $B$  can be positive, negative, or zero in the activate-deactivate phase (depending on  $K$  and  $C_{ON}$ ).

Intuitively, increasing the value of  $\Theta$ , and thus the value of  $I$  (by Equation 4.2), corresponds to larger clusters of ON-cells and OFF-cells forming, which would in turn decrease the pseudo-energy since the first term in Equation 4.3 equals  $-\alpha N\Theta$ . Despite these similarities, we emphasize that the cellular lattice is not the same as an Ising spin

system. For one, there is no real Hamiltonian in our framework that, for instance, gives rise to a Boltzmann distribution. Importantly, we have not used any quantities from physics in our framework, despite some similar properties shared by the framework presented here and those of statistical physics. In the discussion section at the end, we will elaborate further on these similarities and differences.



**Figure 4.3: A cellular lattice acts as a particle that rolls down on and adheres to a pseudo-energy landscape.** (A) Pseudo-energy landscape with a height defined by the pseudo-energy function  $h(p, l)$ . Orange ball is a particle that represents a cellular lattice. The landscape is defined over a position  $(p, l)$ . A pseudo-energy landscape for (B) activate phase, (C) deactivate phase, and (D) activate-deactivate phase. (B-D) Trajectories of the same color start from the same position in each landscape. Black curves show maximally allowed value of the spatial index  $l$  (i.e., function defined as  $l_{\max}(p)$  in the main text; see Supplemental Information section S4.5.2).

#### 4.2.6. GRADIENT OF THE PSEUDO-ENERGY AND THE TRAPPING PROBABILITY $P_{eq}(p, l)$ COMPLETELY SPECIFY THE PARTICLE'S MOTION

To deduce how exactly the shape of the pseudo-energy landscape determines the particle's motion, we compared the gradient field of the pseudo-energy  $-\nabla h(p, l)$  (arrows in Figures 4.4A–D) with the particle trajectories produced by the cellular automaton (red curves in Figures 4.4A–D). We discovered that the particles closely follow the streamlines that are dictated by the gradient field. From this and the aforementioned observation that the particles move stochastically, we conjectured that the particles may follow Langevin-type dynamics in which the particle drifts (rolls) down the pseudo-energy

landscape due to the gradient field and diffuses due to a noise term. We then proposed a phenomenological equation of motion for the particle,

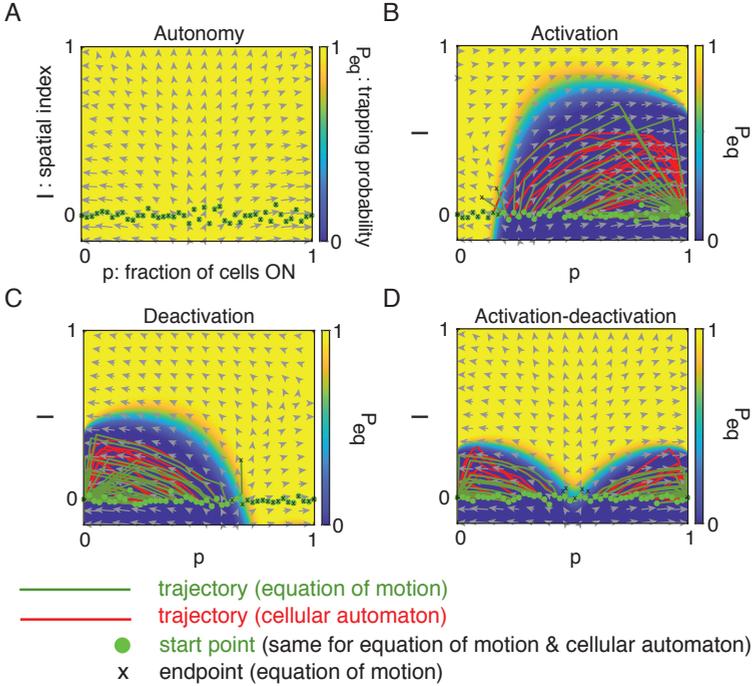
$$(\Delta p(t), \Delta I(t)) = -\nabla h(p(t), I(t)) \cdot \delta + (\eta_p(t), \eta_I(t)) \quad (4.4)$$

Here  $\Delta p(t)$  and  $\Delta I(t)$  are changes in  $p$  and  $I$  respectively between time steps  $t$  and  $t + 1$ ,  $\delta$  is a constant factor that scales the gradient to account for the discreteness of time in the cellular automaton, and  $\eta_p$  and  $\eta_I$  are Gaussian noise terms that represent our ignorance of the microstates with a mean of zero and standard deviations of  $\sigma_p$  and  $\sigma_I$  respectively. We determined  $\delta$ ,  $\sigma_p$  and  $\sigma_I$  by calculating the mean and the variance of  $\Delta p$ , which in turn are set by the distribution of the signal concentrations that each cell senses for a given  $(p, I)$  (Supplemental Information sections S4.5.5–S4.5.6).

While the pseudo-energy determines the direction and the magnitude of changes in  $p$  and  $I$ , it does not predict where a particle stops on the landscape. As we noted earlier (Observation (A)), the particle can stop before its value of  $p$  reaches zero or one. This corresponds to stopping at inclined regions of the pseudo-energy landscape. For this reason, we consider the landscape to be “adhesive”, such that the particle can stop moving on its inclined regions. The gradient of the pseudo-energy is non-zero at such inclined locations, but the particle stops because it has adhered to the landscape at that location. Such particle-adhesions occur frequently for the activate-deactivate phase and in the autonomy phase (e.g. termination points of the brown trajectories in Figure 4.3D). Crucially, the particle halts in a stochastic manner, meaning that for two particles that pass through the same location  $(p, I)$ , one may get stuck there while the other does not. This is because each macrostate  $(p, I)$  can include microstates that are steady states of the cellular automaton and microstates that are not. We need a probabilistic description of how likely it is that a particle at a given location halts since we do not know which microstate is represented by the moving particle when we run a Monte Carlo simulation of Equation 4.4. To obtain a stochastic description, we used a mean-field approach to estimate, for a given macrostate  $(p, I)$ , the fraction of microstates in it that are steady states of the cellular automaton (Supplemental Information section S4.5.5). We call this fraction, which is between zero and one, a “trapping probability” and denote it by  $P_{eq}(p, I)$ . It is the probability that a particle at location  $(p, I)$  corresponds to a steady-state of the cellular automaton and thus halts there. Roughly speaking, the trapping probability  $P_{eq}(p, I)$  represents the “adhesiveness” of the landscape that we discussed earlier.

To produce particle trajectories, we ran a Monte Carlo simulation that combines the phenomenological equation of motion (Equation 4.4) and the condition that the particle halts at location  $(p, I)$  with a probability  $P_{eq}(p, I)$  (Supplemental Information section S4.5.6). We found that the particle trajectories obtained from these Monte Carlo simu-

lations (green curves in Figures 4.4A–D) recapitulated, for a wide range of parameters, the main qualitative features of the particle trajectories that the cellular automaton produces (red curves in Figures 4.4A–D), including the general regions where particles gets stuck, despite some deficiencies (Figures S4.2–S4.3). We will discuss the limitations of this approach in the discussion section at the end.



**Figure 4.4: Gradient field of the pseudo-energy,  $-\nabla h(p, l)$ , and the trapping probability,  $P_{eq}(p, l)$ , together specify the particle's trajectory.** (A–D) Each grey arrow represents the negative gradient of the pseudo-energy,  $-\nabla h(p, l)$ , at each position  $(p, l)$ . Longer arrows indicate gradients of larger magnitudes. Heat maps show the magnitude of the trapping probability  $P_{eq}$  at each location (Supplemental Information section S4.5.5). Red trajectories are exact particle trajectories from the cellular automaton. Green trajectories are particle trajectories produced by Monte Carlo simulations that are dictated by the equation of motion (Equation 4.4) and the trapping probability. The green dots represent the starting points of the trajectories (same for the trajectories produced by the cellular automaton and the equation of motion), and the black crosses represent the endpoints of the green trajectories. (A) Autonomy phase, (B) activate phase, (C) deactivate phase, and (D) activate-deactivate phase. See also Figures S4.2–S4.3.

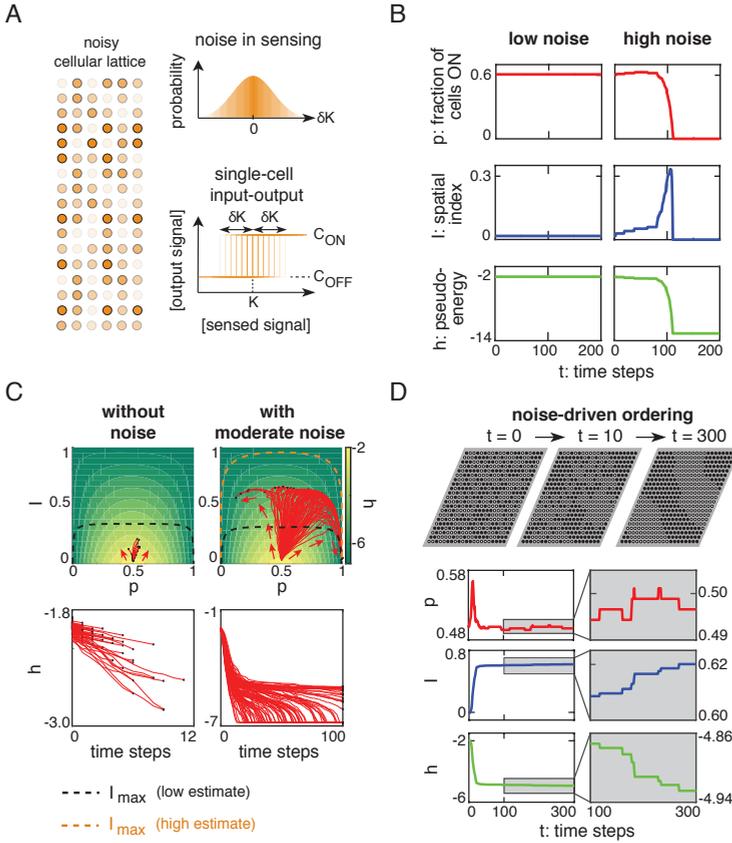
#### 4.2.7. STOCHASTIC SENSING CAN YIELD SPATIAL CONFIGURATIONS THAT ARE MORE ORDERED THAN THOSE FORMED WITHOUT NOISE.

Having shown where the particle gets stuck on the pseudo-energy landscape, a natural question is how stably the particle sticks at each location. Biological noise is a sensible context to address this question. To address this question and as a proof-of-principle for

demonstrating how to include stochastic gene-expression in our framework [Raj & Van Oudenaarden, 2008; Sagues et al., 2007; Garcia-Ojalvo, 2011; Tkačik & Walczak, 2011; Sanchez & Golding, 2013; Xu et al., 2016; Friedman et al., 2006], we modified the deterministic cellular automaton that we have been using thus far to include stochastic sensing. Specifically, for each cell and at each time step of the cellular automaton, we now pick a new value for the activation threshold,  $K + \delta K$ . Here,  $K$  is the same value for every cell at all times and  $\delta K$  is a Gaussian noise term with a mean of zero and a variance of  $\alpha^2$  (Figure 4.5A).

We then define a “noise strength”,  $\xi = \alpha/K$ , that helps us determine how much noise is required to liberate an adhered particle and cause a moving particle to significantly deviate from the path that it would have taken if there were no noise. Intuitively, we would expect such deviations to occur if the noise  $\delta K$  is sufficiently large, such that either an ON-cell, on which the average signal-concentration  $\langle Y_{ON} \rangle$  is larger than the activation threshold without the noise,  $K$ , would turn off due to the noise increasing the activation threshold so that it becomes larger than  $\langle Y_{ON} \rangle$ , or an OFF-cell, on which the average signal-concentration  $\langle Y_{OFF} \rangle$  is smaller than  $K$ , would turn on due to the noise decreasing the activation threshold so that becomes smaller than  $\langle Y_{OFF} \rangle$ . Mathematically, this means that we would expect the minimum noise strength  $\xi_{min}$  required to significantly perturb the particle trajectories to be  $\min(|\langle Y_{ON} \rangle - K|, |\langle Y_{OFF} \rangle - K|) / (K\sqrt{N})$ . Indeed, we found that a very weak noise (i.e.,  $\xi \ll \xi_{min}$ ) cannot detach an adhered particle from the landscape (Figure 4.5B – left column) while a very strong noise (i.e.,  $\xi \gg \xi_{min}$ ) can detach an adhered particle, and thereby cause the particle to roll down the landscape further. After being detached, the particle further changes its  $p$ , decreases its pseudo-energy, and increases its spatial index until its  $p$  reaches either zero or one (Figure 4.5B – right column).

Moreover, we found that a moderate noise (i.e.,  $\xi \sim \xi_{min}$ ) can liberate the adhered particle and push it further down the landscape, beyond the previously allowed region of the landscape (i.e., beyond the region bounded by  $I_{max}(p)$  (black curve in Figure 4.5C)), until it adheres to the landscape again, but now with a higher spatial index than before and with an intermediate value of  $p$  (i.e.,  $0 < p < 1$ ) (Figure 4.5C). Intriguingly, when there is a moderate noise in the activate-deactivate phase, we observed that some of the trapped particles’  $p$ ,  $I$ , and  $h$  very slowly changed over time, allowing the particles to remain stuck with an intermediate value of  $p$  over hundreds but not thousands of time steps (Figure 4.5D).



**Figure 4.5: Stochastic sensing can yield spatial configurations that are more ordered than those formed without noise.** (A) (Left column) Schematics of secrete-and-sense cells with noisy sensing. Each cell (circle) is colored by a different shade of orange, with a darker shade representing less noise. (Top right panel) Noise in activation threshold  $K$ , denoted  $\delta K$ , is normally distributed with a zero mean and a variance  $\alpha^2$ . (Bottom right panel) Range of activation thresholds  $K + \delta K$  for each cell. (B) Examples of changing fraction  $p$  of cells that are ON, spatial index  $I$ , and pseudo-energy  $h$  for low noise (left column;  $\xi < \xi_{min}$ ) and high noise (right column;  $\xi > \xi_{min}$ ) in the activate-deactivate phase.  $\xi = \alpha/K$  is the noise strength and  $\xi_{min}$  is the minimum noise strength required to detach an adhered particle. Both the low noise and the high noise scenarios begin with a spatial configuration that is a steady state of the deterministic cellular automaton. (C) Particle trajectories (red curves), in activate-deactivate phase, for a deterministic cellular automaton (left column) and cellular automaton with a moderate noise (i.e.,  $\xi < \xi_{min}$ ) in sensing (right column). All trajectories start at  $(p = 0.5, I \approx 0)$ . Black dots show endpoints of trajectories. Maximum  $I$  as a function of  $p$  calculated through two distinct approximate schemes. The lower approximation (black curve) corresponds to a maximum in the case without noise and the higher approximation (orange curve) when a moderate noise is present (also see Section S4.5.2). (D) (Top panel) Snapshots at different times of cellular lattice becoming more ordered due to noise in sensing in the activate-deactivate phase. Black circles are ON-cells and white circles are OFF-cells. (Bottom panel): Fraction  $p$  of cells that are ON (red curve), spatial index  $I$  (blue curve), and pseudo-energy  $h$  (green curve) over time for the pattern formation shown in the top panel. Zoomed-in views (grey boxes) show slowly changing  $p$  (red curve),  $I$  (blue curve), and  $h$  (green curve) that occur while the cellular lattice is in a highly ordered metastable configuration (shown at  $t = 300$  in the top panel).

### 4.3. DISCUSSION

Here we have uncovered a visual landscape describing how communicating cells form patterns. Specifically, we considered a ubiquitous form of cellular communication, called secrete-and-sensing. We showed how the landscape determines how secrete-and-sense cells' gene expressions become more spatially correlated over time in the absence of any pre-existing morphogens. Instead of focusing on how specific spatial patterns such as stripes and islands emerge, we focused on the overall spatial order — by characterizing it through a statistical measure of cell-cell coordination of gene expressions that we called spatial index. This macrostate-level description has the advantage of making exhaustive, numerical simulations that are typically used for these systems unnecessary but has the disadvantage of being ignorant of the specific spatial patterns that form. The spatial index, however, still allows us to discern what kinds of spatial patterns are formed because fixing its value restricts the spatial patterns that are possible (Figure 4.2A). Despite its wide applicability, there are instances where the current framework would not apply. We now turn to discussing these situations before concluding with a discussion on how our framework is distinct from those of physics and how one can apply our model to experiments.

4

Our modeling framework for secrete-and-sense cells with bistable (ON/OFF) gene-expression relied on meeting two conditions: (1) Every cell adjusting its ON/OFF-state within the same timescale, and (2) The concentration of the signaling molecule on each cell reaching a steady state before the cell can switch its ON/OFF-state. The first condition sets the actual time that each discrete time-step of the cellular automaton represents and is the reason that the cellular automaton simultaneously updated every cell's state. It is satisfied if the variability among cells in their response times to the signaling molecule (i.e., time taken by each cell to change between ON- and OFF-state) is smaller than the average response-time of the cell. The second condition, which states that the typical response time of the cells is larger than the time that the signaling molecule takes to form a steady-state concentration, is satisfied in several biological processes. They include the aforementioned yeasts that secrete-and-sense the mating pheromone and the regenerating hair follicles in mice [Youk & Lim, 2014a; Chen et al., 2015; Rappaport & Barkai, 2012; Maire & Youk, 2015b]. The condition is also satisfied by several quorum-sensing bacteria (e.g., 20-30 seconds to establish a steady-state concentration) [Kaplan & Greenberg, 1985; Pearson et al., 1999].

Despite these examples, a major aspect that we have neglected is that signaling molecules are often affected by processes other than diffusion such as active transporting of the molecules, and clustering and endocytosis of receptors. Several studies of morphogen gradients in developing embryos, however, have shown that in many cases, one can use

a simple diffusion alone to mathematically reproduce the creation dynamics of morphogen gradients even when there are other processes [Lander et al., 2002]. Finally, aside from conditions (1) and (2), our model assumes that cells are arranged on a triangular lattice. Indeed, several systems, including the nuclei inside the early *Drosophila melanogaster* embryo, can be approximated as being arranged on a triangular lattice despite not satisfying both conditions (1) and (2) [Gregor et al., 2007b] (see other examples in Table S4.1). For other regular lattices, one can modify the framework by changing the functional form of the interaction strength  $f_N(a_0)$ .

Another element in our framework whose validity requires a careful thought is the equation of motion (Equation 4.4). The equation of motion is a phenomenological equation that recapitulates the main qualitative features of the particle trajectories but does not reproduce the exact location of the particle at every time step of the cellular automaton (Figures 4.4A–D). For example, given any initial value of the fraction  $p$  of ON-cells, the equation of motion accurately predicts whether the  $p$  will increase, decrease, or stay the same. But the trajectories produced by the equation of motion do not exactly match those produced by the cellular automaton. In particular, the trajectories produced by the equation of motion are least likely to match those of the cellular automaton at locations where the gradient vector of the pseudo-energy is perfectly horizontal (i.e., parallel to the  $p$ -axis) or vertical (i.e., parallel to the  $I$ -axis), and most likely to match when the gradient is at  $45^\circ$  with respect to both axes (see Section S4.5.6). Since the gradient is neither perfectly horizontal nor vertical (Figures 4.4A–D) at most locations, the gradient of the pseudo-energy together with the trapping probability  $P_{eq}(p, I)$  gives a qualitatively accurate description of the particle's motion.

We also found that the equation of motion gives a more accurate description of the particle trajectories for strong interactions (i.e.,  $f_N(a_0) > 1$ ) than weak interactions (i.e.,  $f_N(a_0) < 1$ ). To see why this is, note that we used mean-field approximations, in which we assumed that ON- and OFF-cells are randomly distributed, to determine the values of  $\sigma_p$ ,  $\sigma_I$  and  $\delta$  in the equation of motion (Equation 4.4) (Supplemental Information section S4.5.6). This mean-field approximation breaks down if long-lived, large islands of ON- and OFF-cells form and slowly grow over time. Such islands indeed frequently form when the interaction is weak and lead to the cellular automaton producing higher values of the spatial index  $I$  than the equation of motion allows for (Figure S4.2). In contrast, when the interaction is strong, the particle typically moves faster because the effect of changing the ON/OFF-state of a single cell propagates to the far-away cells. Thus, the entire lattice of cells typically turns on or off in a few time steps without clearly forming local domains of ON/OFF-cells that grow over time (Figure S4.3). Hence the equation of motion is more suitable for strong interactions than weak interactions.

Finally, we note that another source of quantitative disagreements between the equation of motion and the cellular automaton lies in the fact that in computing the gradient of the pseudo-energy, the equation of motion assumes that  $p$  and  $I$  are continuous variable when in fact they are discrete quantities since the number of cells  $N$  is finite. This continuum approximation, however, is valid in the limit of the population size approaching infinity. This is because the spacing between two adjacent values of  $p$  is  $1/N$  and the spacing between two adjacent values of  $I$  for a fixed value of  $p$  scales as  $1/N$  (see Section S4.5.1).

## 4

In this chapter, we have shown that it is possible build a physics-type framework for complex multicellular systems that are governed by chemical signals and gene-regulatory networks. Many such systems are currently only treated by exhaustive, numerical simulations and are lacking analytic frameworks of the type that we presented here. This situation has risen because the established metrics from physics, such as energy and momentum, are ill-suited for describing gene expressions and chemical signals in multicellular systems. While researchers have used physics-type frameworks to explain many-body living systems such as birds that flock together [Vicsek et al., 1995] and tissues that are subject to mechanical forces [Graner and Glazier, 1992], multicellular systems of the type that we studied here, which are not governed by mechanical or electrical means, have been difficult to treat by directly applying existing concepts and quantities from physics. Despite the similarities in the approach that we have taken and that of statistical mechanics, our framework should not be interpreted in terms of existing quantities from physics because our model does not use any existing quantities of physics such as energy, force, momentum, or temperature. For example, the pseudo-energy (Equation 4.3) only mathematically takes the same form as the long-ranged Ising Hamiltonian. However, the particle does not follow the equations of Hamiltonian mechanics. As another example, the concepts of detailed balance and thermal equilibrium do not apply to the particle that is stuck on the pseudo-energy landscape. In other words, there is no state in which the macroscopic variables remain constant while the cellular lattice dynamically transitions between microstates of the same macrostate. The notions of entropy and temperature also do not have straightforward definitions in our system. One can count the total number of microstates for a given  $(p, I)$  or the number of steady-state microstates for a given  $(K, C_{ON})$  [Maire & Youk, 2015a], but neither would be a thermodynamic entropy. In light of these considerations, it would be interesting to explore, in a future work, if the quantities in our framework can be derived from the quantities of physics.

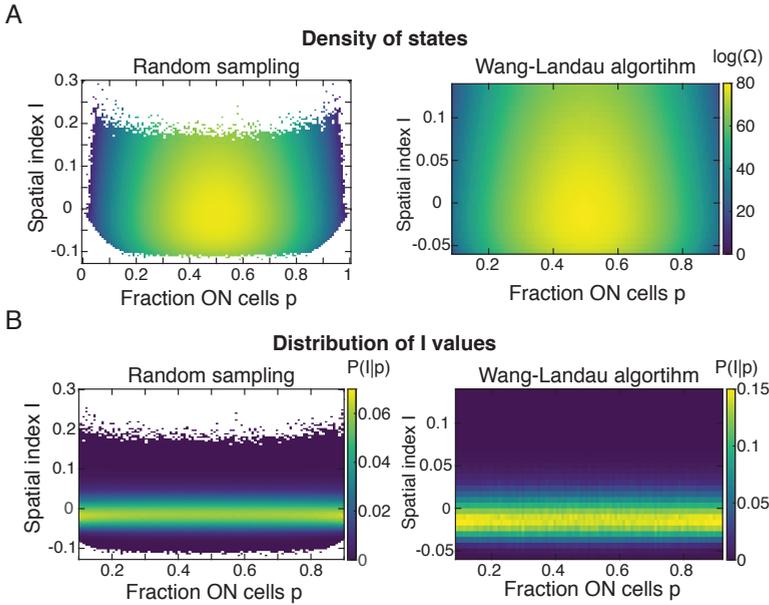
Experimentally, one can measure the two macrostate variables,  $p$  and  $I$ , in microscope images (e.g., by tagging fluorescent protein(s) to the output gene(s)). One may also use

the tools of optogenetics to engineer the cells so that shining light on a single cell would cause the cell to secrete a signaling molecule or switch between the ON- and OFF-state [Guglielmi et al., 2016]. One can then use light to sculpt a pattern of secreting ON-cells at the beginning of an experiment, in effect initializing the values of  $p$  and  $I$ , and then observe how the ON- and OFF-states change by recording over time the fluorescence of each cell that reports whether the cell is ON or OFF. Our model and its extensions may help in understanding such microscope-based time-lapse movies of secrete-and-sense cells that form spatial patterns. Along with studying how specific spatial patterns, such as stripes and islands, are generated, it is useful to focus on statistically describing how certain classes of spatial patterns arise without knowing the exact spatial patterns involved as we have done here. This is because one often cannot measure all the parameters that are required for constructing detailed numerical models (e.g., gene-expression level of every cell in a tissue). In such situations, our framework and its extensions may help in predicting, based on a limited knowledge of the underlying gene-regulation scheme and an estimate of the system's initial spatial order, how the spatial configuration of the cells evolves over time without revealing the exact location, shape, and size of the resulting spatial pattern. We hope that our work, along with complementary approaches for studying spatial patterns [Cotterell et al., 2015; Surkova et al., 2009; Sokolowski et al., 2012; Tkačik et al., 2008; Hillenbrand et al., 2016; Erdmann et al., 2009; Fancher & Mugler, 2017; Thalmeier et al., 2016], will inform on-going efforts to establish quantitative frameworks for multicellular gene-regulations.

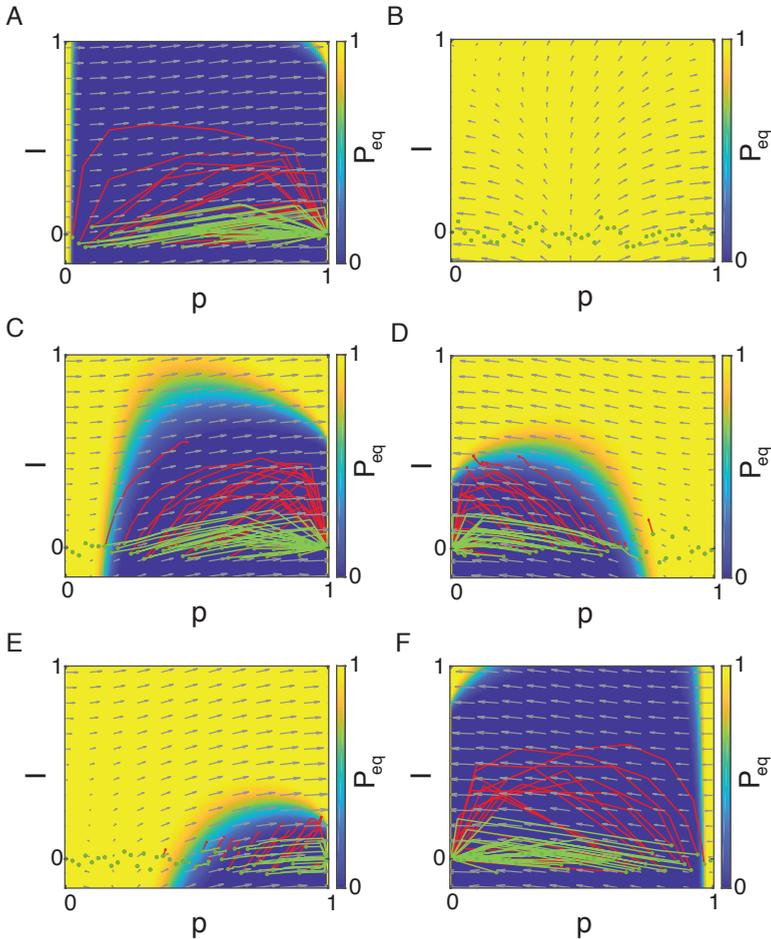
## S4.4. SUPPLEMENTAL DATA ITEMS

Biological System	$a_0$	$R$	$\lambda$	$\tau_{ss}$	References (see caption)
Hair follicle regeneration	150 $\mu\text{m}$	25-50 $\mu\text{m}^*$	1mm	<1 day	[1]
<i>Drosophila melanogaster</i> Bicoid-Hunchback	8.5 $\mu\text{m}$	3.25 $\mu\text{m}$	100 $\mu\text{m}$	90 min	[2]
<i>Drosophila melanogaster</i> ommatidia	17.5 $\mu\text{m}$	8.75 $\mu\text{m}^{**}$	- $^\dagger$	- $^\dagger$	[3-6]
<i>Drosophila melanogaster</i> wing development	3 $\mu\text{m}$	1.45 $\mu\text{m}$	220 $\mu\text{m}$	6-8 hours	[7-9]
Zebrafish embryogenesis Nodal activator	20 $\mu\text{m}$	10 $\mu\text{m}$	135 $\mu\text{m}$	- $^\dagger$	[10]
<i>Xenopus laevis</i> growth factor	20 – 30 $\mu\text{m}$	10 – 15 $\mu\text{m}$	100 $\mu\text{m}^{***}$	5 min $^{****}$	[11, 12]

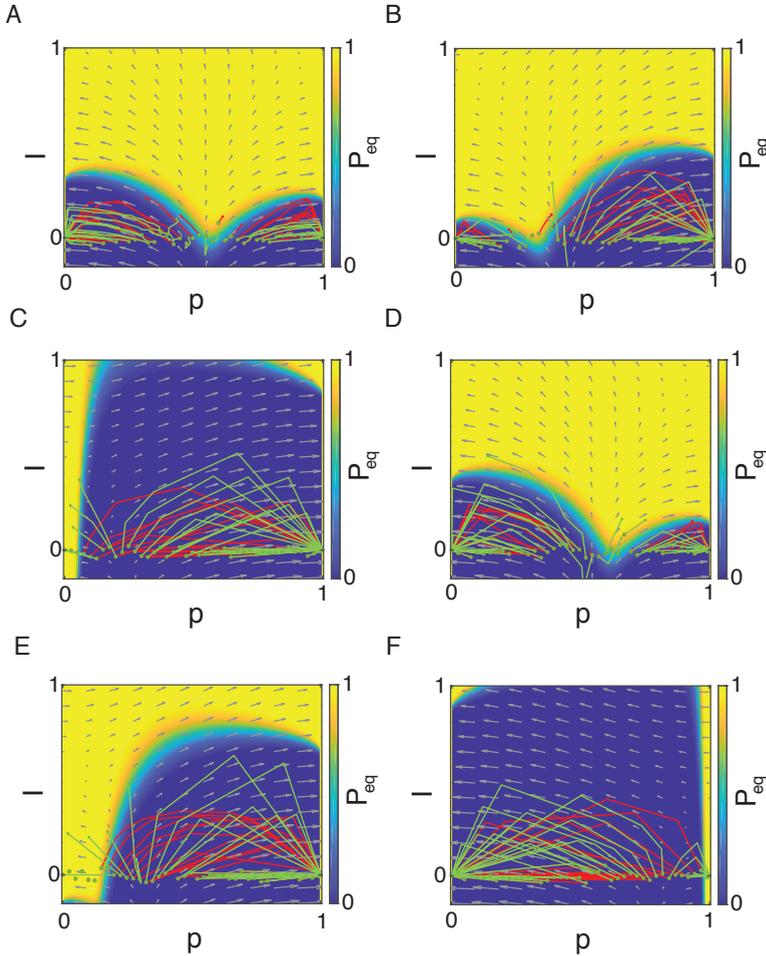
**Table S4.1:** (Related to Figure 4.1) **Secrete-and-sense systems that are arranged on nearly triangular lattices that motivated our work.**  $a_0$  is the approximate distance between the centers of the cells.  $R$  is the average radius of the cells. The signaling molecule has a characteristic diffusion length  $\lambda$ .  $\tau_{ss}$  is the time needed to establish a steady state profile. *Notes:*  $^*$ The radius, measured as the typical size of the hair follicle's base, was inferred from the images in the reference.  $^{**}$ The structure is an extruded hexagon and cannot be approximated by a single radius. The reported value is half the distance between the centers of the ommatidia units.  $^{***}$  Assumed to be half the length of the gradient.  $^{****}$   $\tau_{ss}$  not reported in paper, but calculated from the estimate  $\tau_{ss} \sim r^2/D$  where  $r$  is the typical signaling length in the system and  $D$  is the diffusion constant (reported).  $^\dagger$ Quantity not known or not measured. References in the table: [1] Chen et al., *Cell* (2015); [2] Gregor et al., *Cell* (2007b); [3] Mikeladze-Dvali et al., *Cell* (2005); [4] Posnien et al., *PLoS ONE* (2012); [5] Tsachaki and Sprecher, *Dev. Dynamics* (2012); [6] Yang et al. *Cell* (2002); [7] Entchev et al., *Cell* (2000); [8] Lander et al., *Dev. Cell* (2002); [9] Teleman and Cohen, *Cell* (2000); [10] Müller et al., *Science* (2012); [11] Green, *Dev. Dynamics* (2002); [12] McDowell et al., *Int. J. Dev. Biol.* (2001)



**Figure S4.1:** (Related to Figure 4.2) **Density of states, indicating the number of microstates (lattice configurations) belonging to the same macrostate ( $p, I$ ).** We used two methods of calculating this density of states (see Section S4.5.3). In random sampling, we generated a large set of random lattice configurations and bin together configurations with the same values of  $p$  and  $I$  to get statistics. With the Wang-Landau algorithm, we efficiently sample over a pre-defined region in  $p - I$  space using a Monte Carlo algorithm. (A) Overall density of states  $\Omega$  determined by random sampling (left) and the Wang-Landau algorithm (right). (B) Distribution of values of  $I$  given a fixed value of  $p$ , i.e. the conditional distribution  $P(I|p)$ , for the same two methods. Note that the axes in (A) and (B) have different scales.



**Figure S4.2:** (Related to Figure 4.4) **Gradient fields of the pseudo-energy,  $P_{eq}$ , and particle trajectories — Weak interactions.** Simulated trajectories of the cellular automaton (red lines), together with trajectories of the equation of motion Eq. S4.41 (green lines), plotted on top of the vector field generated by the negative of the gradient of the pseudo-energy (grey arrows), plotted on top of the ‘stickiness’  $P_{eq}$  (color bar). Note that  $P_{eq} = 1$  at  $p = 0$  and  $p = 1$  for all these examples (sometimes marked by a thin yellow line near the edge). Circles represent initial values and crosses values at equilibrium. For the Langevin trajectories (green), we took the same initial values as generated from the automaton simulations. Weak interaction regime ( $a_0 = 1.5$ ),  $N = 121$  cells. (a) Activation,  $K = 3, C_{ON} = 24$ , (b) Autonomy,  $K = 15, C_{ON} = 20$ , (c) Activation,  $K = 6, C_{ON} = 21$ , (d) Deactivation,  $K = 17, C_{ON} = 14$ , (e) Activation,  $K = 10, C_{ON} = 21$ , (f) Deactivation,  $K = 20, C_{ON} = 14$ .



**Figure S4.3:** (Related to Figure 4.4) **Gradient fields of the pseudo-energy,  $P_{eq}$ , and particle trajectories — Strong interactions.** Same as Figure S4.2, but now for the strong interaction regime ( $a_0 = 0.5$ ) and  $N = 121$  cells. (a) Activation-deactivation,  $K = 10, C_{ON} = 5$ , (b) Activation-deactivation,  $K = 19, C_{ON} = 14$ , (c) Activation,  $K = 10, C_{ON} = 21$ , (d) Activation-deactivation,  $K = 16, C_{ON} = 8$ , (e) Activation,  $K = 14, C_{ON} = 16$ , (f) Deactivation,  $K = 18, C_{ON} = 6$ .

## S4.5. SUPPLEMENTARY INFORMATION

In the Supplementary Information section, we provide details of methods, derivations and calculations related to the main text of the chapter. We have opted to include a shortened version of the original Supplementary Information section published in [Olimpio et al., 2018]. Specifically, we leave out the sections on the model extensions that include noise and multiple cell types altogether. The main text still contains a figure and section on lattices with noise, but leaves out multiple cell types. Furthermore, we have replaced two more sections — the derivation of  $I_{max}$  (Section S4.5.2) and the proof that  $h$  is a non-increasing function of time (Section S4.5.4) — by concise summaries outlining the main results and sketches of the derivations. These results originated from the master thesis of Eduardo Olimpio [Olimpio, 2016], but are not essential for understanding the rest of the chapter. On the other hand, we have also included one more section which was not featured in our publication [Olimpio et al., 2018], which deals with the density of states as a function of  $(p, I)$  (Section S4.5.3). This work originated from reviewer comments that we addressed in a rebuttal, but did not include in the final publication.

### S4.5.1. PROPERTIES OF THE SPATIAL INDEX $I$

DERIVATION OF THE SPATIAL INDEX  $I$  IN TERMS OF  $p$

The **spatial index**,  $I$  is a modified version of the Moran index (i.e., Moran's  $I$ ). It is a weighted, spatial autocorrelation of the cell states whereby each cell pair  $(i, j)$  is weighted by the interaction strength  $f_N(r_{ij})$  for that pair. Specifically, we defined  $I$  in the main text as

$$I = \frac{N}{\sum_i \sum_{j \neq i} f(r_{ij})} \frac{\sum_i \sum_{j \neq i} f(r_{ij}) (X_i - \langle X \rangle) (X_j - \langle X \rangle)}{\sum_i (X_i - \langle X \rangle)^2}, \quad (\text{S4.1})$$

where  $f(r_{ij})$  is the term in the interaction strength for the cell-pair  $(i, j)$ :  $f(r_{ij}) = \frac{e^{R-r_{ij}}}{r_{ij}} \sinh(R)$  (this is denoted by  $g(r_{ij})$  in the main text). Moreover,  $X_i$  is defined by

$$X_i = \begin{cases} -1 & \text{if cell-}i \text{ is OFF} \\ 1 & \text{if cell-}i \text{ is ON} \end{cases}, \quad (\text{S4.2})$$

which is different from the previous definition of  $X_i$  (Chapter 3.1). In the remainder of this supplementary text, we will use this definition of  $X_i$  for reasons that will become clearer in the following sections. Note that this definition does not change the features of our model because our previous definition of  $X_i$  is related to this revised definition of  $X_i$  by a linear relation ( $2X_i^{\text{old}} = X_i^{\text{new}} + 1$ ).

In terms of  $p$ , we have  $X_i$  is  $\langle X \rangle = 2p - 1$  and  $\sum_i (X_i - \langle X \rangle)^2 = 4Np(1 - p)$ . Moreover,

by the definition of the interaction strength (equation 3.5), we have  $\sum_{j \neq i} f(r_{ij}) = f_N$ . Therefore

$$I = \frac{1}{f_N} \frac{\sum_i \sum_{j \neq i} f(r_{ij}) X_i X_j - 2(2p-1) \sum_i \sum_{j \neq i} f(r_{ij}) X_i + (2p-1)^2 N f_N}{4Np(1-p)}$$

where we used the fact that  $f(r_{ij}) = f(r_{ji})$ . Note that

$$\sum_i \sum_{j \neq i} f(r_{ij}) X_i = N \langle \sum_{j \neq i} f(r_{ij}) X_i \rangle = N f_N \langle X \rangle = N f_N (2p-1)$$

where the brackets denote averaging among all cells. Combining above results, we have

$$I = \frac{\langle \sum_{j \neq i} f(r_{ij}) X_i X_j \rangle - (2p-1)^2 f_N}{4p(1-p) f_N}, \quad (\text{S4.3})$$

where we have used  $\sum_i \sum_{j \neq i} f(r_{ij}) X_i X_j = N \langle \sum_{j \neq i} f(r_{ij}) X_i X_j \rangle$ . Note that  $f_N$  is a purely geometric quantity that is almost constant when  $N$  is sufficiently large due to  $f(r_{ij})$  scaling as  $e^{-r_{ij}}$ . This is precisely equation 2 in the main text. For later use, we also define

$$\Theta = \frac{1}{N} \sum_i \sum_{j \neq i} f(r_{ij}) X_i X_j. \quad (\text{S4.4})$$

With this, we can write

$$I = \frac{\Theta(X) - (2p-1)^2 f_N}{4p(1-p) f_N}. \quad (\text{S4.5})$$

SPATIAL INDEX  $I$  IN THE LIMITS  $p \rightarrow 0$  AND  $p \rightarrow 1$

Let  $X = (X_1, \dots, X_N)$  be a microstate, with  $X_i \in \{-1, 1\}$ . Note that  $I$  is undefined when  $p = 0$  or  $p = 1$  because both denominator and nominator vanish in Equation S4.1. So we can only discuss what  $I$  becomes in the limit of  $p \rightarrow 0$  and  $p \rightarrow 1$ . Let us consider the limit  $p \rightarrow 0$  (same argument applies to  $p \rightarrow 1$ ). Given that  $p$  is a discrete variable for a fixed value of  $N$ , taking the limit  $p \rightarrow 0$  means that we consider the value of  $I$  at  $p = 1/N$ , which is the lowest possible non-zero value of  $p$ . For a lattice with  $N$  cells, there are  $N$  microstates with  $p = 1/N$ . In these states, all but one cell is OFF. Let  $X_1 = 1, X_{i \neq 1} = -1$ , then

$$\begin{aligned} \Theta(X) &= \frac{1}{N} \sum_{j \neq 1} f(r_{1j}) X_1 X_j + \frac{1}{N} \sum_{i=2}^N \sum_{\substack{j=2 \\ j \neq i}}^N f(r_{ij}) X_i X_j = \frac{1}{N} (-f_N + \sum_{i=2}^N \sum_{\substack{j=2 \\ j \neq i}}^N f(r_{ij})) \\ &= \frac{1}{N} (-f_N + \sum_{i=2}^N (f_N - f(r_{i1}))) = \frac{1}{N} (-f_N + (N-2) - f_N) = \frac{N-4}{N} f_N. \end{aligned} \quad (\text{S4.6})$$

Also,  $p = 1/N$  and thus

$$I = \frac{\frac{N-4}{N} f_N - \frac{(N-2)^2}{N^2} f_N}{\frac{4(N-1)}{N^2} f_N} = \frac{(N-4)N - (N-2)^2}{4(N-1)} = -\frac{1}{N-1}. \quad (\text{S4.7})$$

For typical lattice sizes that we study (e.g.,  $N = 225$ ), the above equation tells us that  $I \sim -0.001$ , which is practically zero. Moreover, we see that for  $N \rightarrow \infty$ ,  $I \rightarrow 0$ . In fact, for  $N \rightarrow \infty$ , we also have  $p \rightarrow 0$  for this configuration. For these reasons, we set  $I = 0$  when  $p \rightarrow 0$  in our study. The same holds for the limit  $p \rightarrow 1$ . Thus, defining  $I = 0$  for a uniform lattice seems to be a reasonable choice and ensures continuity in the limit of  $N \rightarrow \infty$ .

4

#### UPPER BOUND FOR THE SPACING BETWEEN ALLOWED VALUES OF $I$

To justify the fact that in the macroscopic equation of motion (Section S4.5.6) we take  $(p, I)$  to be continuous, we show in this section that the spacing between possible values of  $I$  is bound by a value that goes to zero in the limit  $N \rightarrow \infty$ . Note that the spacing in allowed values of  $p$  is  $1/N$  and therefore trivially goes to zero. Determining the spectrum of allowed values of  $I$  for a fixed  $p$  is a notably harder problem, which we will not tackle in detail here. Rather, we will only derive an upper bound that goes to zero in the large system size limit, for any  $p$  not too close to zero or one (more on this later).

Consider a microstate with  $p \neq 0$ ,  $p \neq 1$  and consider two cells  $k$  and  $l$ . Suppose  $X_k \neq X_l$ , i.e. one of them is an ON-cell and the other an OFF-cell. We will consider what happens to  $I$  if we flip both cells, i.e.  $X_k \rightarrow -X_k$  and  $X_l \rightarrow -X_l$  (here we take  $X_i \in \{-1, 1\}$ ). Clearly, the fraction of ON-cells does not change, so  $p$  remains constant. For the change in  $\Theta$ , let us first write

$$N\Delta\Theta = \sum_{i \neq k, l} \sum_{j \neq i, k, l} f(r_{ij}) X_i X_j + 2 \sum_{j \neq k} f(r_{kj}) X_k X_j + 2 \sum_{j \neq l} f(r_{lj}) X_l X_j - 2f(r_{kl}) X_k X_l. \quad (\text{S4.8})$$

The factors 2 come from the fact that each interaction is counted twice in the definition of  $\Theta$  (Equation S4.4). Here we have separated the terms of  $\Theta$  into four terms, of which only the middle two change when we flip the states of cells  $k$  and  $l$ . Hence we have

$$N\Delta\Theta \equiv N(\Theta_{new} - \Theta_{old}) = -4 \sum_{j \neq k} f(r_{kj}) X_k X_j - 4 \sum_{j \neq l} f(r_{lj}) X_l X_j. \quad (\text{S4.9})$$

Since  $|X_i| = 1$  for all cells and  $|\sum_{j \neq k} f(r_{kj})| \leq \sum_{j \neq k} |f(r_{kj})| = f_N$ , we obtain

$$\begin{aligned} |N\Delta\Theta| &\leq 4 \left| \sum_{j \neq k} f(r_{kj}) X_k X_j \right| + 4 \left| \sum_{j \neq l} f(r_{lj}) X_l X_j \right| \leq 8 \left| \sum_{j \neq k} f(r_{kj}) X_k X_j \right| \\ &\leq 8 \sum_{j \neq k} |f(r_{kj}) X_k X_j| \leq 8 \sum_{j \neq k} |f(r_{kj})| |X_k| |X_j| = 8 f_N. \end{aligned} \quad (\text{S4.10})$$

Finally, let  $\Delta I = I_{new} - I_{old}$ . Since  $p$  does not change, we have

$$|\Delta I| = \left| \frac{\Delta\Theta}{4f_N p(1-p)} \right| \leq \frac{2}{Np(1-p)}. \quad (\text{S4.11})$$

This calculation shows that starting from an arbitrary lattice in which not all cells are ON or OFF, it is always possible to generate a different cellular lattice with the same  $p$ , of which the value of  $I$  differs by no more than  $\frac{2}{Np(1-p)}$ . In the limit of  $N \rightarrow \infty$ , this value goes to zero whenever  $p$  is not too close to 0 or 1. Therefore, as long as we are away from the boundaries, we can safely take  $I$  to be continuous in the limit of large system size.

For  $p$  close to zero and one, the bound might be very large (but note that the above argument excludes the extremes  $p = 0$  and  $p = 1$ ). However, there are very few different values of  $I$  that are possible near these bounds. For instance, for a single ON-cell in a lattice of OFF cells there is only one value of  $I$  possible. For two ON-cells, the number of unique values equals the number unique distances possible between two cells. As we argued in the main text and in the previous section, the value of  $I$  becomes irrelevant in these limits as there are only one or a few values possible. Nevertheless, we should be careful when considering the value of  $I$  near these bounds.

### S4.5.2. DERIVING THE MAXIMUM ALLOWED VALUE OF $|I|$ FOR EACH $p$

*Detailed results of this section are presented in Supplementary Information section S3 of Olimpio et al., 2018. Here we give a summary of the main results.*

We derived two approaches for estimating the maximum allowed values for  $I$  as a function of  $p$ . The first approach relies on a mean-field approximation and results in the dashed black curves in Figs. 4.3 and 4.5. We first estimate  $I$  through a mean-field approximation, where we exactly calculate the contributions to the sensed concentrations from nearest neighbors, and estimate the contributions from the rest of the system. Then, we express the result as a function of the average number of ON nearest neighbors  $\langle m_{i=ON} \rangle$ . This quantity  $\langle m_{i=ON} \rangle$  in turn depends on the geometry of the clusters of ON-cells and can be fully expressed in terms of the areas and perimeters of connected domains (polygons) of ON-cells. This idea is analogous to polygon constructions used to study the

critical point in the Ising model. Based on these two results, one can then argue that the geometry that maximizes the spatial index  $I$  is when all ON-cells are packing into a single polygon. Since the same argument can be made for OFF-cells, we obtain two bounds  $I_{\max}^{ON}(p)$  and  $I_{\max}^{OFF}(p)$ . We then choose  $I_{\max}(p) = \max(I_{\max}^{OFF}(p), I_{\max}^{ON}(p))$ . Explicitly, the result is

$$I_{\max}^{ON} \approx \left( 6(1-p) - \frac{6}{pN} - \frac{8}{\sqrt{pN}} \right) \frac{f(a_0)}{(1-p)f_N}, \quad (\text{S4.12})$$

$$I_{\max}^{OFF} \approx \left( 6p - \frac{6}{(1-p)N} - \frac{8}{\sqrt{(1-p)N}} \right) \frac{f(a_0)}{pf_N}. \quad (\text{S4.13})$$

4

This first method slightly underestimates the true maximum allowed values of  $I$  (dashed black curves in Figs. 4.3 and 4.5).

The second construction relies on making a continuum approximation for the sum over all cells and results in the dashed orange curve in Figure 4.5C. Specifically, we replace the sum by an integral weighted by a “density of states”, which is the density of cells at a given distance  $r$  from an arbitrary cell. This is comparable to the density of states (in terms of energy) in condensed matter physics used to infer e.g. electronic properties of solids. One then again estimates the maximum by assuming that all cells of the same state (ON or OFF) are clustered together in a single (spherical) domain. Next, by using the expression for  $I$  in terms of  $p$  (without  $\Theta$ ; Equation S4.3), one obtains the corresponding value of  $I$ , which can be expressed as:

$$I_{\max} = \frac{f_N - 2G(pe^{-L_{ON}} + (1-p)e^{-L_{OFF}} - e^{-L}) - (2p-1)^2 f_N}{4p(1-p)f_N}, \quad (\text{S4.14})$$

$$G = 2\pi e^R \sinh(R),$$

$$L_{OFF} = \sqrt{\frac{((1-p)N - 7)A_{\text{cell}}}{\pi} + (1.5a_0)^2},$$

$$L_{ON} = \sqrt{\frac{(pN - 7)A_{\text{cell}}}{\pi} + (1.5a_0)^2}.$$

This construction assumes a perfect separation between ON and OFF cells into spherical domains, and further makes a continuum approximation for the density of states. As a result, the value of  $I_{\max}$  is overestimated (dashed orange curve in Figure 4.5C).

### S4.5.3. DENSITY OF STATES AS A FUNCTION OF $(p, I)$

This this section, we discuss two methods to calculate the density of states, i.e. how many microstates fit into the same macrostate  $(p, I)$ . Denote  $\Omega(p)$  as the number of states with

a fixed value of  $p = k/N$ ,  $k \in 0, 1, \dots, N$  – this is easily obtained as the binomial  $\binom{N}{k}$ . Define  $\Omega(p, I; N, a_0)$  as the total number of states given  $(p, I)$ . Let us also define  $P(p, I) = \frac{\Omega(p, I)}{2^N}$  as the fraction of states for given  $(p, I)$ . The calculation of these quantities is non-trivial due to way  $I$  is defined. Below we represent two methods of computing approximate values of these quantities as functions of  $(p, I)$ .

#### RANDOM SAMPLING

Naively, the most straightforward method of estimation is by randomly sampling over a large number of configurations. Concretely, one generates random configurations, computes the value of  $(p, I)$  for each configuration, and then bins the sampled results to obtain  $P(p, I)$ . As the number of states  $2^N$  is typically much larger than the sample size, we cannot expect it to give highly accurate results. In particular, rare events that occur with low probability are likely to be missed, so that the sampled distribution tends to become strictly zero beyond a certain region, while the actual distribution is not.

One slight improvement to fully random sampling is what we will call ‘conditional sampling’. In this case, we fix one of our variables at a time and then sample over configurations with that fixed variable to obtain the conditional distribution of the other variable. In our system, it is straightforward to fix  $p$  and generate random configurations to obtain the conditional distribution  $P(I|p) = P(p, I)/P(p)$ . Since  $P(p) = \binom{N}{Np}$ , we obtain  $P(p, I)$  and  $\Omega(p, I)$  by computing  $P(I|p)$  for each value of  $p$ .

#### WANG-LANDAU ALGORITHM

A method that is tailored to obtain a better estimate for especially the outlier values of  $I$  is the Wang-Landau (WL) algorithm [Wang & Landau, 2001]. This is mainly used to calculate the density of states (DOS) — i.e. the number of states at a fixed value of energy — of spin systems such as the Ising model. The main idea is to perform a random walk in phase space by proposing a spin flip at each step, that is accepted with a probability similar to the Metropolis algorithm, but where less frequent states are *more* likely to be accepted. Starting from an initial guess, we gradually update a function that gradually becomes closer and closer to the DOS up to a proportionality factor.

We exploit the similarity of our system to a spin model and adapt the WL algorithm to calculate  $P(I|p)$ , which we will informally refer to as our DOS. The first main difference is that the random walk needs to be over states with fixed  $p$ . This is achieved by flipping two cell states at a time, one random OFF cell and a random ON cell. Also, the ‘density’ is for the spatial order parameter  $I$  rather than some energy function. We thus obtain the following detailed scheme (adapted and modified from [Barros et al., 2015]):

1. Start with an arbitrary configuration of with fixed  $p$  and a guess for the DOS, which we take to be  $g(I) = 1$  for all possible  $I$ .

2. Choose an OFF cell and an ON cell at random and make the configuration with the selected OFF cell turned ON and the ON cell turned OFF. Compute the spatial order parameter for the original configuration,  $I_1$ , and for the proposed new configuration,  $I_2$ . Accept the change with probability

$$p(I_1 \rightarrow I_2) = \min\left(\frac{g(I_1)}{g(I_2)}, 1\right). \quad (\text{S4.15})$$

This ensures that changes to states that have been less visited (for which  $g(I_2) < g(I_1)$ ) are always accepted.

3. Let  $I$  be the resulting value of the trial flip. We then update our function by a modification factor  $f > 1$ ,

$$g(I) \rightarrow fg(I). \quad (\text{S4.16})$$

4. The resulting  $I$  value is stored in a histogram (with bins that need to be manually specified).

$$H(I) \rightarrow H(I) + 1. \quad (\text{S4.17})$$

5. Repeat steps 2-4 until the histogram  $H(I)$  is sufficiently flat. This is imposed through a flatness parameter  $p_{\text{flat}} < 1$ , which specifies how close all bins have to be to the average bin value  $\langle H(I) \rangle$ . The condition then becomes

$$H(I) > p_{\text{flat}} \times \langle H(I) \rangle \quad (\text{S4.18})$$

for all bins of the histogram.

6. After the flatness criterion is reached, we update the modification factor  $f$  through

$$f \rightarrow (f)^{\frac{1}{2}}. \quad (\text{S4.19})$$

If  $f < f_{\text{final}}$ , we are done.

7. Otherwise, we reset the histogram,  $H(I) \rightarrow 0$  for all  $I$ , and repeat the entire procedure above with the new value of  $f$ .

In the implementation we work with logarithms of  $g(I)$ , because the values quickly become too large to handle for a typical simulation. The parameters we need to set in the simulation are  $f$ ,  $f_0$ ,  $f_{\text{final}}$ , the histogram bins for  $I$ , and  $p_{\text{flat}}$ . We take the histogram to have 30 bins over the interval  $[-0.06, 0.14]$ . Choosing a larger value leads to longer running times as it becomes harder to sample over rare values of  $I$  (or an infinite loop if one of the bins is fully outside the permitted range of  $I$ ). We take the rest of the pa-

rameters to be similar to typical values suggested for the Ising model [Wang & Landau, 2001]:  $f = f_0 = e \approx 2.71828$ ,  $p_{\text{flat}} = 0.8$ . However, we take  $f_{\text{final}} = \exp(10^{-4})$ , which leads our simulation to finish in less time than the suggested  $f_{\text{final}} = \exp(10^{-8})$ . Finally, we normalize by dividing our result for  $g(I)$  by  $\sum_{I \in \text{bins}} g(I)$ . In practice we typically need to rescale  $g(I)$  to avoid overflow problems, for instance by dividing all values by  $\min_{I \in \text{bins}} g(I)$ . Note that this does not influence the result since  $g(I)$  is proportional to  $P(I|p)$ .

#### S4.5.4. PROOF THAT THE PSEUDO-ENERGY $h$ IS A NON-INCREASING FUNCTION OVER TIME

*Detailed results of this section are presented in Supplementary Information section S4 of Olimpico et al., 2018. Here we give a summary of the main results.*

4

We first express the change in  $H$  at any given time as a function of the changes in the cell states  $\Delta X_m$  (notation:  $\Delta X_m(t) = X_m(t+1) - X_m(t)$ , i.e. the state change for cell  $m$ ):

$$\begin{aligned} \Delta H &= -2 \sum_m \Delta X_m (Y_m - K) - B \sum_m \Delta X_m, \\ B &= \left( \frac{C_{ON} + 1}{2} (1 + f_N) - K \right). \end{aligned} \quad (\text{S4.20})$$

One can then argue that  $2 \sum_m \Delta X_m (Y_m - K) \geq 0$  and analyze the behavior for the different behavioral phases individually (cf. Section 4.2.2 in the main text). It is then straightforward to show that  $\Delta H \leq 0$  in the (1) the autonomy phase, for which  $\Delta H = 0$  by definition, (2) activation phase, for which  $B > 0$  and  $\sum_m \Delta X_m \geq 0$ , and the (3) deactivation phase, for which  $B < 0$  and  $\sum_m \Delta X_m \leq 0$ . The only non-trivial case to prove is then the activation-deactivation phase. The proof here relies on first proving that  $\Theta$  monotonically increases over time, and then expressing  $h$  explicitly in terms of  $p$  and  $I$ . Explicitly, we obtained

$$\begin{aligned} h(p, I) &= -\frac{(C_{ON} - 1)}{2} (1 + 4f_N p(1-p)I + (2p-1)^2 f_N), \\ &\quad - (2p-1) \left[ \frac{(C_{ON} + 1)}{2} (1 + f_N) - K \right]. \end{aligned} \quad (\text{S4.21})$$

Through Equation S4.5, one can then express  $h$  in terms of  $(p, \Theta)$ . Then  $\Theta \geq 0$  implies  $\Delta h \leq 0$  whenever two conditions for  $\Delta p$  in terms of the signal field  $B$  are satisfied. We empirically verified these conditions in simulations.

### S4.5.5. DERIVATION OF THE TRAPPING PROBABILITY — $P_{eq}(p, I)$

The concentration  $Y_i$  sensed by cell- $i$  is:

$$Y_i = Y_i^{\text{self}} + Y_i^{\text{nei}}, \quad (\text{S4.22})$$

where  $Y_i^{\text{self}}$  is the self-contribution (i.e., signal secreted by cell- $i$  itself) and  $Y_i^{\text{nei}}$  is the contribution from all other cells. Specifically, they are

$$Y_i^{\text{self}} = (C_{ON} - 1) X_i + 1, \quad (\text{S4.23a})$$

$$Y_i^{\text{nei}} = \sum_{j \neq i} f(r_{ij}) [(C_{ON} - 1) X_j + 1]. \quad (\text{S4.23b})$$

The probability that an ON-cell remains ON in the next time step is the same as the probability  $P_{ON \rightarrow ON}$  that the signal concentration on an ON-cell is larger than the threshold concentration  $K$ :

$$P_{ON \rightarrow ON} = P(Y_i > K | X_i = 1) = P(Y_i^{\text{nei}} > K - C_{ON} | X_i = 1). \quad (\text{S4.24})$$

Similarly defining  $P_{OFF \rightarrow OFF}$  as the probability that an OFF cell senses a concentration of the signal that is lower than  $K$ , we have

$$P_{OFF \rightarrow OFF} = P(Y_i < K | X_i = 0) = P(Y_i^{\text{nei}} < K - 1 | X_i = 0). \quad (\text{S4.25})$$

We will show below that  $P_{ON \rightarrow ON}$  and  $P_{OFF \rightarrow OFF}$  are expressible in terms of  $p$  and  $I$ . If we randomly pick a microstate out of a "box" that belongs to a macrostate  $(p, I)$ , the probability that the microstate is an equilibrium state is given by the **trapping probability**  $P_{eq}(p, I)$ , where

$$P_{eq}(p, I) = (P_{ON \rightarrow ON})^n (P_{OFF \rightarrow OFF})^{N-n}. \quad (\text{S4.26})$$

Since we do not know exactly where each ON- and OFF-cell is but only know that there is a total of  $n$  ON-cells, we treat  $X_i$  for all  $i$  as a random variable. Accordingly,  $Y_i^{\text{nei}}$  is also a random variable. Furthermore, by placing the cells in the lattice in a completely random fashion,  $Y_i^{\text{nei}}$  does not depend on the state of cell- $i$ ,  $X_i$ . Thus

$$P_{ON \rightarrow ON} = P(Y_i^{\text{nei}} > K - C_{ON}), \quad (\text{S4.27a})$$

$$P_{OFF \rightarrow OFF} = P(Y_i^{\text{nei}} < K - 1). \quad (\text{S4.27b})$$

We assume that  $X_i$  follows a binomial distribution with a probability  $p = n/N$ . Then the mean of the  $X_i$  distribution is  $p$  and the variance is  $(1 - p)p$ . For a large  $N$ , the Central

Limit Theorem dictates that  $\langle X_i \rangle$  is normally distributed:

$$\langle X_i \rangle \sim \mathcal{N}(p, p(1-p)), \quad (\text{S4.28})$$

where  $\mathcal{N}(\mu, \sigma^2)$  is the normal distribution with mean  $\mu$  and variance  $\sigma^2$ . To find how  $Y_i^{\text{nei}}$  is distributed, we use equation S4.23b. Since each cell's position is fixed,  $Y_i^{\text{nei}}$  is linear in  $X_i$  and is thus also normally distributed. This means that we only need to compute the mean and the variance of  $Y_i^{\text{nei}}$ , which depend on both  $p$  and the spatial index  $I$ . Because  $I$  is related to whether a cell has neighbors that are similar or different from itself, the sensed concentration for an OFF-cell will be different from for an ON-cell within the same configuration whenever  $I \neq 0$ . For an OFF-cell, we write  $Y_{i=\text{OFF}}^{\text{nei}} \sim \mathcal{N}(\mu_{\text{OFF}}, \sigma_{\text{OFF}})$ , whereas for an ON-cell we have  $Y_{i=\text{ON}}^{\text{nei}} \sim \mathcal{N}(\mu_{\text{ON}}, \sigma_{\text{ON}})$ .

4

To calculate the means and variances, we first change notation and define  $\tilde{X} = 2X - 1$ , so that  $\tilde{X} \in \{-1, 1\}$ . Then we can write

$$Y_i^{\text{nei}} = \left( \frac{C_{\text{ON}} + 1}{2} \right) f_N + \left( \frac{C_{\text{ON}} - 1}{2} \right) \sum_{i \neq j} f(r_{ij}) \tilde{X}_j. \quad (\text{S4.29})$$

Taking the average of above, we have

$$\mu_{\text{ON}} = \left( \frac{C_{\text{ON}} + 1}{2} \right) f_N + \left( \frac{C_{\text{ON}} - 1}{2} \right) \left\langle \sum_{i \neq j} f(r_{ij}) \tilde{X}_j \middle| i = \text{ON} \right\rangle, \quad (\text{S4.30a})$$

$$\mu_{\text{OFF}} = \left( \frac{C_{\text{ON}} + 1}{2} \right) f_N + \left( \frac{C_{\text{ON}} - 1}{2} \right) \left\langle \sum_{i \neq j} f(r_{ij}) \tilde{X}_j \middle| i = \text{OFF} \right\rangle. \quad (\text{S4.30b})$$

Note that we can write the expression for  $\Theta$  (Equation S4.4) as

$$\Theta = p \left\langle \sum_{i \neq j} f(r_{ij}) \tilde{X}_j \middle| i = \text{ON} \right\rangle - (1-p) \left\langle \sum_{i \neq j} f(r_{ij}) \tilde{X}_j \middle| i = \text{OFF} \right\rangle. \quad (\text{S4.31})$$

At the same time, using Bayes' theorem we have

$$\left\langle \sum_{i \neq j} f(r_{ij}) \tilde{X}_j \right\rangle = (2p-1) f_N = p \left\langle \sum_{i \neq j} f(r_{ij}) \tilde{X}_j \middle| i = \text{ON} \right\rangle + (1-p) \left\langle \sum_{i \neq j} f(r_{ij}) \tilde{X}_j \middle| i = \text{OFF} \right\rangle. \quad (\text{S4.32})$$

Combining S4.31 and S4.32 allows us to solve for  $\langle \sum_{i \neq j} f(r_{ij}) \tilde{X}_j \middle| i = \text{ON} \rangle$  and  $\langle \sum_{i \neq j} f(r_{ij}) \tilde{X}_j \middle| i =$

OFF). Combined with S4.30 we obtain

$$\langle Y_{ON}^{\text{nei}} \rangle \equiv \mu_{ON} = f_N [C_{ON}p + 1 - p + (C_{ON} - 1)(1 - p)I], \quad (\text{S4.33a})$$

$$\langle Y_{OFF}^{\text{nei}} \rangle \equiv \mu_{OFF} = f_N [C_{ON}p + 1 - p - (C_{ON} - 1)pI]. \quad (\text{S4.33b})$$

For the variance of the sensed concentration of the signal, we change back to the original notation with  $X_i \in \{0, 1\}$ . From a mean-field approximation we then obtain

$$\langle Y_i^{\text{nei}} \rangle = \sum_{j \neq i} [(C_{ON} - 1) \langle X_j \rangle + 1] \frac{e^{R-r_{ij}}}{r_{ij}} \sinh(R) = [(C_{ON} - 1)p + 1] f_N(a_0) \equiv \mu_p. \quad (\text{S4.34})$$

4

Moreover, we have

$$\langle (Y_i^{\text{nei}})^2 \rangle = \sum_{j \neq i} (C_{ON} - 1)^2 \langle X_j^2 \rangle \frac{e^{2(R-r_{ij})}}{r_{ij}^2} \sinh^2(R) = (C_{ON} - 1)^2 p(1 - p) g_N(a_0) \equiv \sigma_p^2, \quad (\text{S4.35})$$

where we have defined the function

$$g_N(a_0) \equiv \sum_{j \neq i} \frac{e^{2(R-r_{ij})}}{r_{ij}^2} \sinh^2(R). \quad (\text{S4.36})$$

From this, we obtain

$$\sigma_{ON} = \sigma_{OFF} = (C_{ON} - 1) \sqrt{g_N p(1 - p)}. \quad (\text{S4.37})$$

Combining above results, we finally have

$$P_{ON \rightarrow ON} = 1 - D(K - C_{ON}; \mu_{ON}, \sigma_{ON}) = \frac{1}{2} \left[ 1 - \operatorname{erf} \left( \frac{K - C_{ON} - \mu_{ON}}{\sqrt{2} \sigma_{ON}} \right) \right], \quad (\text{S4.38a})$$

$$P_{OFF \rightarrow OFF} = D(K - 1; \mu_{OFF}, \sigma_{OFF}) = \frac{1}{2} \left[ 1 + \operatorname{erf} \left( \frac{K - 1 - \mu_{OFF}}{\sqrt{2} \sigma_{OFF}} \right) \right], \quad (\text{S4.38b})$$

where  $D(x; \mu, \sigma)$  is the cumulative distribution function of the Gaussian with mean  $\mu$  (equation S4.33) and standard deviation  $\sigma$  (equation S4.37) while  $\operatorname{erf}(x)$  is the error function.

#### S4.5.6. EQUATION OF MOTION DERIVED FROM THE PSEUDO-ENERGY

In this section we derive the equation of motion (Equation 4.4), determine the values of the new variables in this equation of motion and discuss on a technical level the validity of the approach. A more colloquial discussion is included in the main text.

From observations, we see that the simulated trajectories follow paths whose directions

tend to point in the direction of fastest descent of the pseudo-energy (Figure 4.4). This motivates us to construct an equation that utilizes the direction of fastest descent. The gradient of the pseudo-energy is given by  $\vec{\nabla}h = (\frac{\partial h}{\partial p}, \frac{\partial h}{\partial I})$ , with

$$\begin{aligned}\frac{\partial h}{\partial p} &= -(C_{ON} - 1)2f_N(1 - I)(2p - 1) - (C_{ON} + 1)(f_N + 1) + 2K, \\ \frac{\partial h}{\partial I} &= -2f_N(C_{ON} - 1)p(1 - p).\end{aligned}\tag{S4.39}$$

Note that the above expression implies that there are no local minima of  $h$  which are not on the boundary of the  $(p, I)$  phase space. This follows from the fact that the gradient must vanish at a local minimum, and the only points at which  $\frac{\partial h}{\partial I}$  vanishes are  $(p, I) = (0, 0)$  and  $(p, I) = (1, 0)$  (taking into account the fact that  $p = 0$  and  $p = 1$  have only one microstate with  $I = 0$ , see Section S4.5.1). Hence for  $0 < p < 1$  there can be no local minima. Therefore, the trapped configurations alluded to in the main text cannot be directly related to the minima of the pseudo-energy.

Recall that the direction of steepest descent at any point  $(p, I)$  is given by the negative of the gradient,  $-\vec{\nabla}h$ . As a first attempt, we can therefore try as phenomenological equations

$$\begin{aligned}\frac{\partial p}{\partial t} &= -\frac{\partial h}{\partial p}, \\ \frac{\partial I}{\partial t} &= -\frac{\partial h}{\partial I}.\end{aligned}\tag{S4.40}$$

While in some cases the streamlines produced by this vector field closely approach the automaton simulations, Equation S4.40 misses out an important aspect of the macrostate-level dynamics. The dynamics it produces is deterministic, providing only one possible trajectory for a given starting macrostate  $(p_0, I_0)$ . This ignores the fact that the macrostate-level description is degenerate — a macrostate  $(p, I)$  usually has many microstates, which could follow multiple non-identical trajectories.

Hence we need to introduce noise terms to represent our ignorance of the microstates. We choose to add white noise, constructing a Langevin equation where the (negative of the) gradient represents the drift term and the Brownian motion corresponds to the variability between trajectories of different initial microstates. The next caveat is that the cellular automaton operates in discrete time, where a Langevin equation consisting of S4.40 with added noise terms would naturally be described in continuous time. To better assess the compatibility of the Langevin equation with our discrete time cellular automaton, we need to modify it into a discrete time equation whose time steps reflect average changes in the automaton.

To do this, we first introduce a step size  $\delta$ , which is a scaling factor that controls how far along the negative of the gradient the system should travel in one time step. Our discrete-time system then becomes

$$\begin{aligned}\Delta p &= -\frac{\partial h}{\partial p}\delta + \eta_p, \\ \Delta I &= -\frac{\partial h}{\partial I}\delta + \eta_I.\end{aligned}\tag{S4.41}$$

Here we have introduced the noise as Gaussian variables  $\eta_p \sim \mathcal{N}(0, \sigma_p)$ ,  $\eta_I \sim \mathcal{N}(0, \sigma_I)$ . Next, we will derive expressions for these parameters based on the microstate-level details of the system.

4

MEAN-FIELD CALCULATION OF  $\Delta p$ , USED FOR OBTAINING  $\delta$

In this section we provide a calculation of  $\Delta p$  based on the switching probabilities  $P_{ON \rightarrow ON}$  and  $P_{OFF \rightarrow OFF}$  used in Section S4.5.5. By comparing the result with Equation S4.41 we can fix the value of  $\delta$ , as we will show in the next section.

Recall that  $\Delta p$  is the amount of change in  $p$  at time step  $t$ . Using the probabilities calculated in the section S4.5.5, we can calculate by how much  $p$  changes at time  $t$ :  $\Delta p_t = p_{t+1} - p_t$ . We will use this to obtain a constant scale factor  $\delta$  that rescales the gradient of the pseudo-energy in the equation of motion. We assume that all ON-cells have a binomial chance of transitioning to an OFF state with probability  $1 - P_{ON \rightarrow ON}$ . Therefore, if we have  $n$  ON-cells, the probability that  $y_-$  ON-cells will switch OFF in the next time step is

$$P(y_-; n, 1 - P_{ON \rightarrow ON}) = P_{ON \rightarrow ON}^{n-y_-} (1 - P_{ON \rightarrow ON})^{y_-} \binom{n}{y_-}.\tag{S4.42}$$

Similarly, the probability that  $y_+$  OFF-cells will switch ON in the next time step is

$$P(y_+; n, 1 - P_{OFF \rightarrow OFF}) = P_{OFF \rightarrow OFF}^{N-n-y_+} (1 - P_{OFF \rightarrow OFF})^{y_+} \binom{N-n}{y_+}.\tag{S4.43}$$

We are now interested in the mean and the variance of  $\Delta p_t$ . Note that

$$\Delta p_t = \frac{1}{N} (y_+ - y_-) \equiv p_+ - p_-.\tag{S4.44}$$

The mean of a binomial distribution with  $N$  draws and probability  $p$  is  $Np$ . Hence, taking the average of equation S4.44, we obtain the average of  $\Delta p_t$ :

$$\langle \Delta p_t \rangle \equiv E(p_+) - E(p_-) = (1-p)(1 - P_{OFF \rightarrow OFF}) - p(1 - P_{ON \rightarrow ON}).\tag{S4.45}$$

To calculate the variance of  $\Delta p_t$ , we assume that we can approximate the distributions S4.42 and S4.43 as being independent Gaussians. This assumption is most accurate when  $n$  and  $N - n$  are both sufficiently large. With this assumption, we obtain the following simplified result for the variance:

$$\begin{aligned} \text{Var}(\Delta p_t) &\approx \text{Var}(p_+) + \text{Var}(p_-), \\ &= \frac{1}{N} [(1-p)(1-P_{OFF \rightarrow OFF})P_{OFF \rightarrow OFF} + p(1-P_{ON \rightarrow ON})P_{ON \rightarrow ON}]. \end{aligned} \quad (\text{S4.46})$$

Note that in the thermodynamic limit ( $N \rightarrow \infty$ ), the variance of  $\Delta p_t$  goes to zero as expected.

#### PARAMETERS OF THE EQUATION OF MOTION

We will use the results derived above to fix  $\delta$ . Recall that the probabilities  $P_{ON \rightarrow ON}$  (Equation S4.24) and  $P_{OFF \rightarrow OFF}$  (Equation S4.25) are functions of  $p$  and  $I$ , so  $\langle \Delta p \rangle$  depends on both  $p$  and  $I$ . This suggests a way of fixing  $\delta = \tilde{\delta}(p)$  defined through

$$\langle \Delta p \rangle = -\frac{\partial h}{\partial p} \tilde{\delta}(p). \quad (\text{S4.47})$$

In this way, the equation of motion would always predict  $\Delta p$  to be equal to  $\langle \Delta p \rangle$  as given in S4.45. However, the quantity  $\tilde{\delta}(p)$  depends on  $p$  through  $\langle \Delta p \rangle$  and  $\frac{\partial h}{\partial p}$  which is inconsistent with how we arrived at S4.41. To obtain a constant scaling factor, we have to average this quantity over a suitable weight function on  $(p, I)$  space. We choose to average this quantity over  $p$  with as weight the fraction of states with  $n$  ON cells or  $p = n/N$ ,

$$f_p = \frac{1}{2^N} \binom{N}{n}. \quad (\text{S4.48})$$

Since  $I$  enters the equations only through  $\frac{\partial h}{\partial p}$ , which depends linearly on  $I$ , we make the approximation that  $I = 0$ . This is justified by calculations in which we find that the density of states (i.e. the number of microstates corresponding to a macrostate  $(p, I)$ ) is highest around  $I \approx 0$  for each value of  $p$  (Figure S4.1). Thus we obtain  $\delta$  through a double averaging procedure as

$$\delta = \sum_{n=0}^N f_p \tilde{\delta}(p) = -\sum_{n=0}^N f_p \langle \Delta p \rangle \left( \frac{\partial h}{\partial p} \right)^{-1}. \quad (\text{S4.49})$$

More explicitly,

$$\delta = \sum_{n=0}^N \frac{1}{2^N} \binom{N}{n} \frac{(1-p)(1-P_{OFF \rightarrow OFF}) - p(1-P_{ON \rightarrow ON})}{(C_{ON} - 1)2f_N(2p - 1) + (C_{ON} + 1)(f_N + 1) - 2K}. \quad (\text{S4.50})$$

Here, we implicitly also take  $I = 0$  in the expressions for  $P_{ON \rightarrow ON}$  and  $P_{OFF \rightarrow OFF}$ .

To estimate the noise, we use S4.45, but now derive

$$\begin{aligned}
 N^2 \text{Var}(\Delta p) &= \text{Var}(y_+) + \text{Var}(y_-) - 2 \text{Cov}(y_+, y_-), \\
 &= \text{Var}(y_+) + \text{Var}(y_-), \\
 &= N(1-p)(1-P_{OFF \rightarrow OFF})P_{OFF \rightarrow OFF}, \\
 &\quad - Np(1-P_{ON \rightarrow ON})P_{ON \rightarrow ON}.
 \end{aligned} \tag{S4.51}$$

The first equality follows from the independence of  $y_+$  and  $y_-$ , and the second follows from the properties of the binomial distribution. As for  $\delta$  we first define a  $p$ -dependent (or  $n$ -dependent) standard deviation

$$\tilde{\sigma}_p(n) = \sqrt{(1-p)(1-P_{OFF \rightarrow OFF})P_{OFF \rightarrow OFF} - p(1-P_{ON \rightarrow ON})P_{ON \rightarrow ON}}. \tag{S4.52}$$

Hence we define the noise in  $\Delta p$  as

$$\sigma_p = \sum_{n=0}^N f_p \tilde{\sigma}_p(n). \tag{S4.53}$$

To obtain the first two moments of  $\Delta I$ , we note that in the absence of noise, any change in  $I$  is related to the change of  $p$  through the gradient (Equation S4.41), namely through

$$\Delta I = \frac{\partial h}{\partial I} \left( \frac{\partial h}{\partial p} \right)^{-1} \Delta p. \tag{S4.54}$$

We now assume that the same relation holds for the first two moments of  $\Delta I$ , giving

$$\begin{aligned}
 \langle \Delta I \rangle &= \frac{\partial h}{\partial I} \left( \frac{\partial h}{\partial p} \right)^{-1} \langle \Delta p \rangle, \\
 \sigma_I &= \frac{\partial h}{\partial I} \left( \frac{\partial h}{\partial p} \right)^{-1} \sigma_p,
 \end{aligned} \tag{S4.55}$$

#### STOPPING CONDITION

Equation S4.41 alone cannot predict where the trajectories will end. If we do not impose any conditions, the trajectories would always escape from the allowed phase space or end at the edge, as we have seen that there can be no points with  $0 < p < 1$  where the gradient vanishes. To fix this problem and predict terminal macrostates for our system, we set up two more rules for the evolution of S4.41. First, let the system terminate at  $(p, I)$  with probability  $P_{eq}(p, I)$ , the trapping probability  $P_{eq}(p, I)$  we derived in Section

S4.5.5. This is done through a Monte Carlo step, in which we draw a random number and compare it with  $P_{eq}(p, I)$  to decide whether to terminate the simulation.

Second, we add an additional rule near the boundaries of the phase space. As the particle trajectories have a tendency to cross  $p = 0$  or  $p = 1$  at values of  $I \neq 0$ , we need to impose an additional stopping condition to prevent the particle from leaving the phase space (i.e.,  $p > 1$  and  $p < 0$ ). Recall that for  $p = 0$  and  $p = 1$  we only have one microstate with  $I = 0$ . Therefore, we set up the additional rule that  $(p, I) \rightarrow (0, 0)$  and  $(p, I) \rightarrow (1, 0)$  when the system attempts to exit the phase space across  $p = 0$  or  $p = 1$  respectively.

#### PARTICLE TRAJECTORIES FROM THE GRADIENT FIELD OF THE PSEUDO-ENERGY AND $P_{eq}$

Given the equation of motion and  $P_{eq}$ , we can predict the particle trajectories and compare them with the particle trajectories produced by the cellular automaton. We find that for all the different behavioural phases, the equation of motion approximates the trajectories well and generally predicts the correct final (resting) configurations (Figs. S4.2, S4.3). A more detailed discussion of these results is given in the main text.

Note that we can clearly identify a region of low  $P_{eq}$  (blue) and a region of high  $P_{eq}$  (yellow) with a relatively sharp transition between them. Furthermore, the cellular automata terminate mostly in regions where  $P_{eq}$  is high. These two features imply that that using only  $P_{eq}$ , we can predict in which regions in  $(p, I)$  the particles will likely come to a rest, if it ever reached the region. Also, it means that there is a graphical way to estimate the particle trajectories and their stopping points in the  $(p, I)$ -space. First, we plot the vector field and  $P_{eq}$  together (Figs. 4 and S4.2-S4.3). From the directions of the vector field, we can estimate how the particle will move. By tracing out these trajectories until the particles reach a region where the  $P_{eq}$  is high, we can get an estimate of the final  $(p, I)$  of the trajectory. As seen in Figs. S4.2-S4.3, this gives a good estimate of the direction and endpoints of the trajectories. The match between the particle trajectories dictated by the equation of motion and those produced by the cellular automaton depends on the parameters (i.e.,  $C_{ON}$ ,  $K$ , and  $f_N$ ), but in general the overall direction and endpoints match well.

#### WHY THE GRADIENT APPROACH WORKS: THE ALLOWED RANGE OF DIRECTIONS

Apart from observing that the particles seem to move in the direction of steepest descent of  $h$ , we can make a more precise mathematical argument as for why  $-\vec{\nabla}h$  is a good estimate for the particle's direction of the motion. The argument that we will give here will also show that the range of allowed directions is much larger (spanning an arc of  $\pi$  radians or half a circle). Whether the trajectories are likely to match with gradient vector field or not in part depends on how its direction relates to this range of allowed directions. In particular, whenever the gradient is (close to) horizontal (i.e., zero degrees)

or vertical (i.e.,  $\pi/2$  radians) in the  $(p, I)$  space, significant deviations from the direction of steepest descent are possible.

The fact that the pseudo-energy is a Lyapunov function means that the system can only move in a direction for which the pseudo-energy decreases. It cannot move in the direction of the (positive of the) gradient  $\vec{\nabla}h$  and by continuity also not in a direction close to it. To be precise, define  $\vec{v} = (\cos\theta, \sin\theta)$  as a unit vector and define  $\theta$  as the angle it makes with the gradient  $\vec{\nabla}h$ . Recall that the system does not have any local minima as shown earlier, so the gradient cannot vanish. Therefore  $\theta$  is well-defined. Define

$$\Delta h(\theta) = \vec{\nabla}h \cdot \vec{v}(\theta). \quad (\text{S4.56})$$

It follows that  $\Delta h(0) > 0$  and  $\Delta h(\pi) < 0$ , as the gradient and negative of the gradient point in directions of steepest ascent and descent respectively. By continuity, there exist  $0 < \theta_1 < \pi$  and  $\pi < \theta_2 < 2\pi$  for which  $\Delta h(\theta_1) = 0$  and  $\Delta h(\theta_2) = 0$  (Intermediate Value Theorem). Hence we can define an interval  $I_\theta \equiv [\theta_1, \theta_2]$  that gives the range of directions in which  $h$  decreases, i.e. for all  $\theta \in I_\theta$ , we have  $\Delta h(\theta) \leq 0$ . From  $\Delta h(\theta) = 0$ , we see that  $\theta_1$  and  $\theta_2$  are solutions to

$$\tan\theta = -\frac{\partial h}{\partial p} \left( \frac{\partial h}{\partial I} \right)^{-1}. \quad (\text{S4.57})$$

This has an exact solution on  $S^1 = \{0 \leq \theta \leq 2\pi | 0 \equiv 2\pi\}$  with  $\theta_2 = \theta_1 + \pi$ . Therefore,  $|I_\theta| = \pi$ . Conversely, the minimum and maximum of  $\Delta h(\theta)$  are defined through  $\frac{d\Delta h(\theta)}{d\theta} = 0$ , for which the solution is  $\tan\theta = \frac{\partial h}{\partial p} \left( \frac{\partial h}{\partial I} \right)^{-1}$ . From the properties of the tangent function, one can check that the solutions to  $\tan x = c$  and  $\tan x = -c$  are close to each other whenever  $|c|$  is either very small or very large. Note that the angles are defined on a circle, so distances are measured by taking the solutions modulo  $2\pi$ . This implies that the minimum is close to the bounds of  $I_\theta$  whenever  $\frac{\partial h}{\partial p} \left( \frac{\partial h}{\partial I} \right)^{-1}$  goes to 0 or  $\infty$ . This can only happen if either  $\frac{\partial h}{\partial p} \rightarrow 0$  or  $\frac{\partial h}{\partial I} \rightarrow 0$ .

#### LIMITATIONS OF THE GRADIENT APPROACH

The argument above shows that the negative of the gradient gives a direction of the particle's motion indication and represents a first-order approximation. However, the range of directions in which the system is in principle able to move (given its macrostate) is in fact much wider — the unit vectors of the allowed directions lie on a half circle which contains  $-\hat{\nabla}h$  (unit vector pointing in direction of  $-\vec{\nabla}h$ ). Whether the trajectories will seem close to the negative of the gradient depends on the position of  $-\hat{\nabla}h$  on this half circle. If  $-\hat{\nabla}h$  points at the middle of this 'allowed arc', then the trajectories can deviate at most a right angle from  $-\hat{\nabla}h$ . However, if  $-\hat{\nabla}h$  is in a direction close to one of the edges, then deviations can in principle approach a limiting value of 180 degrees. We

have shown in the previous section that the latter occurs only if the vector field points close to vertical or horizontal. Indeed, a closer look at the trajectories of Figs. S4.2 – S4.3 reveals that significant deviations from the automaton trajectories are almost exclusively found in cases where the vector field is close to horizontal or vertical. This is apparent in all of the weak interaction pictures (Figure S4.2), but also in the region in which the horizontal component of the vector field changes sign in the strong interaction regime (Figure S4.3).

# 5

## CELLULAR DIALOGUES THAT ENABLE SELF-ORGANIZATION OF DYNAMIC SPATIAL PATTERNS

*Cells form spatial patterns by coordinating their gene expressions. How a group of mesoscopic numbers (hundreds-to-thousands) of cells, without pre-existing morphogen gradients and spatial organization, self-organizes spatial patterns remains poorly understood. Of particular importance are dynamic spatial patterns such as spiral waves that perpetually move and transmit information. We developed an open-source software for simulating a field of cells that communicate by secreting any number of molecules. With this software and a theory, we identified all possible “cellular dialogues” — ways of communicating with two diffusing molecules — that yield diverse dynamic spatial patterns. These patterns emerge despite widely varying responses of cells to the molecules, gene-expression noise, spatial arrangements, and cell movements. A three-stage, “order-fluctuate-settle” process forms dynamic spatial patterns: cells form long-lived whirlpools of wavelets that, following erratic dynamics, settle into a dynamic spatial pattern. Our work helps in iden-*

*tifying gene-regulatory networks that underlie dynamic pattern formations.*

## 5.1. INTRODUCTION

Spatial patterns can form when multiple cells, without pre-existing morphogen gradients, communicate with each other to coordinate their gene expressions [Gregor et al., 2010; Lubensky et al., 2011; Sgro et al., 2013; Idema et al., 2013, Manyukan et al., 2017; Jörg et al., 2019]. Understanding how cells collectively organize spatial patterns through cell-cell communication is crucial for understanding and engineering mammalian tissues [Javaherian et al., 2013]. Many synthetic and natural mammalian tissues are monolayers of genetically identical cells (e.g., epithelial sheets) whose gene-expression levels are initially uncorrelated but become more correlated over time during development, leading to specialized cell types within tissues. This process often involves cell-cell communication [Menendez et al., 2010]. There has been a rising interest in developing experimental methods for spatially arranging individual cells in a monolayer and then observing how such a heterogeneous tissue — composed of cells at differing locations having different gene-expression levels — develops over time [Javaherian et al., 2014]. Although there are quantitative models to explain such experiments, they are often tailored to specific tissues and signaling molecules. Thus, it is challenging to use them as a general framework that one can adapt to different gene-circuits, signaling molecules, and cell types [Drasdo et al., 2007]. Currently unknown is a comprehensive set of generally applicable, quantitative mechanisms by which organized spatial patterns can form in heterogeneous tissues made of mesoscopic numbers (hundreds to thousands) of cells without pre-existing morphogen gradients (Figure 5.1A – top).

To explain pattern formations, one often uses reaction-diffusion equations and Turing instability, in which a uniformly spread field of chemicals develops minute fluctuations in its chemical concentrations at some locations that grow over time to yield spatial patterns (Figure 5.1A – bottom) [Turing, 1952]. Although theoretical studies of Turing instability uncovered many insights into how continuous fields of chemicals or cells form patterns, the instability does not treat gene expressions of individual cells when there are biologically realistic, mesoscopic numbers of cells (Figure 5.1A – top). Furthermore, while

---

An adapted version of this chapter has been published as Y. Dang, D.A.J. Grundel and H. Youk, “Cellular Dialogs: cell-cell communication through diffusible molecules yields dynamic spatial patterns”, *Cell Systems* **10**, 82–98.e7 (2020). Note that the published version includes supplementary videos omitted from the written thesis.

Some results from this chapter have also appeared in the master thesis of D.A.J. Grundel [Grundel, 2019].

Full source code of all scripts used to obtain the results in this chapter is available at: [https://github.com/YitengDang/Cell\\_Systems\\_2019](https://github.com/YitengDang/Cell_Systems_2019).

Full data sets containing all analyzed data used in the figures in this chapter is available at Dryad: <https://doi.org/10.5061/dryad.6hdr7sqw5>

many gene networks can use Turing instability to generate spatial patterns, they are not robust as their circuit parameters need to be finely tuned [Marcon et al., 2016, Scholes et al., 2019]. In light of these difficulties, a promising route for explaining multicellular patterning would be to develop multiscale models that link intracellular signaling with cell-cell communication for mesoscopic numbers of cells. While researchers have developed such models for specific systems — examples include studies of how eyes form [Lubensky et al., 2011] and neurons differentiate [Jörg et al., 2019] — we currently lack a general framework for identifying widely applicable principles of pattern formation. Motivated by this shortcoming, we sought to build a generalized framework that uncovers relationships between properties of cellular communication — the various ways in which the cells secrete and sense signaling molecules — and gene-expression patterns (spatial patterns) that emerge for mesoscopic populations of cells.

## 5

Here we developed an open-source software that simulates spatial-patterning dynamics for a system of communicating cells<sup>1</sup>. One can easily modify and expand our software with more ingredients, and use it for both research and educational purposes. We also developed algorithms for analyzing these simulations. With the software and analysis algorithms, we sought to quantitatively reveal mechanisms by which mesoscopic numbers of cells can form spatial patterns. We focused on dynamic patterns — patterns that constantly change over time without ever stopping such as oscillations and spiral waves [Sgro et al., 2013] — instead of static patterns that remain still after forming (Figure 5.1B). Our computational search discovered all the ways in which cells can communicate with just two diffusing molecules to form dynamic patterns, including those that have been experimentally observed. We found that a few ways of communicating, which we refer to as “cellular dialogues”, can generate a large palette of complex, dynamic spatial patterns such as whirlpools of wavelets and traveling waves of various shapes and orientations. We devised an analytical (pen-and-paper) approach that recapitulates the simulations and used it to understand why only certain cellular dialogues can sustain dynamic spatial patterns. We found that cells form dynamic spatial patterns through a three-stage, “order-fluctuate-settle” process. Starting from a configuration in which there is no spatial correlation among cells’ gene-expression levels, cells rapidly become more spatially correlated over time, resulting in self-organized wavelets. This is followed by a prolonged transient phase in which the wavelets constantly and erratically form and annihilate each other. Finally, as the wavelets settle down, a dynamic spatial pattern such as a traveling wave emerges. We show that self-organized dynamic patterns can still form despite widely varying gene-expression noise, cellular responses to the sensed molecules, spatial arrangements of cells, and diffusive (random) motions of cells. As a

---

<sup>1</sup><https://github.com/YitengDang/MultiCellSim>

theoretical study, we focused on exploring how cells *can* form dynamic spatial patterns, rather than explaining any specific biological system. But our computational screen still uncovered cellular dialogues that are known to generate dynamic spatial patterns in specific multicellular systems. Our paper ends by suggesting how one can expand our work, including the open-source software, to identify as-yet-unknown cellular dialogues that produce known dynamic spatial patterns in multicellular systems.

## 5.2. RESULTS

### 5.2.1. COMPUTATIONAL SEARCH FOR CELLULAR DIALOGUES THAT ENABLE SELF-ORGANIZED PATTERNS

We built a visualization software that simulates all possible ways in which cells can communicate — which we call “cellular dialogues” — by secreting, sensing, and responding to two diffusing molecules (Figure 5.1C). Such cells, which simultaneously secrete and sense one or more signaling molecules, are ubiquitous in nature [Hart et al. 2014; Youk and Lim, 2014a; Youk and Lim, 2014b; Chen et al. 2015; Maire and Youk, 2015b]. Our simulations combine reaction-diffusion equations – describing the concentrations of the molecules – and a cellular automaton – describing the cells’ gene-expression levels that are set by the concentrations of the two molecules. We represent a cellular dialogue as a network diagram that consists of two nodes (one for each molecule) joined by signed arrows, which can be positive (activating) or negative (repressing). A signed arrow denotes how the sensing of one molecule, represented by the node on which the arrow begins, increases (for a positive arrow) or decreases (for a negative arrow) the sensing cell’s secretion rate of a molecule that is represented by the node on which the arrow ends (Figure 5.1C). We assume that both molecules diffuse on a faster timescale than the cells can respond — the two molecules “rapidly” diffuse and reach steady-state concentrations to which the cells then respond — as is the case in many multicellular systems [Heemskerk et al., 2019].

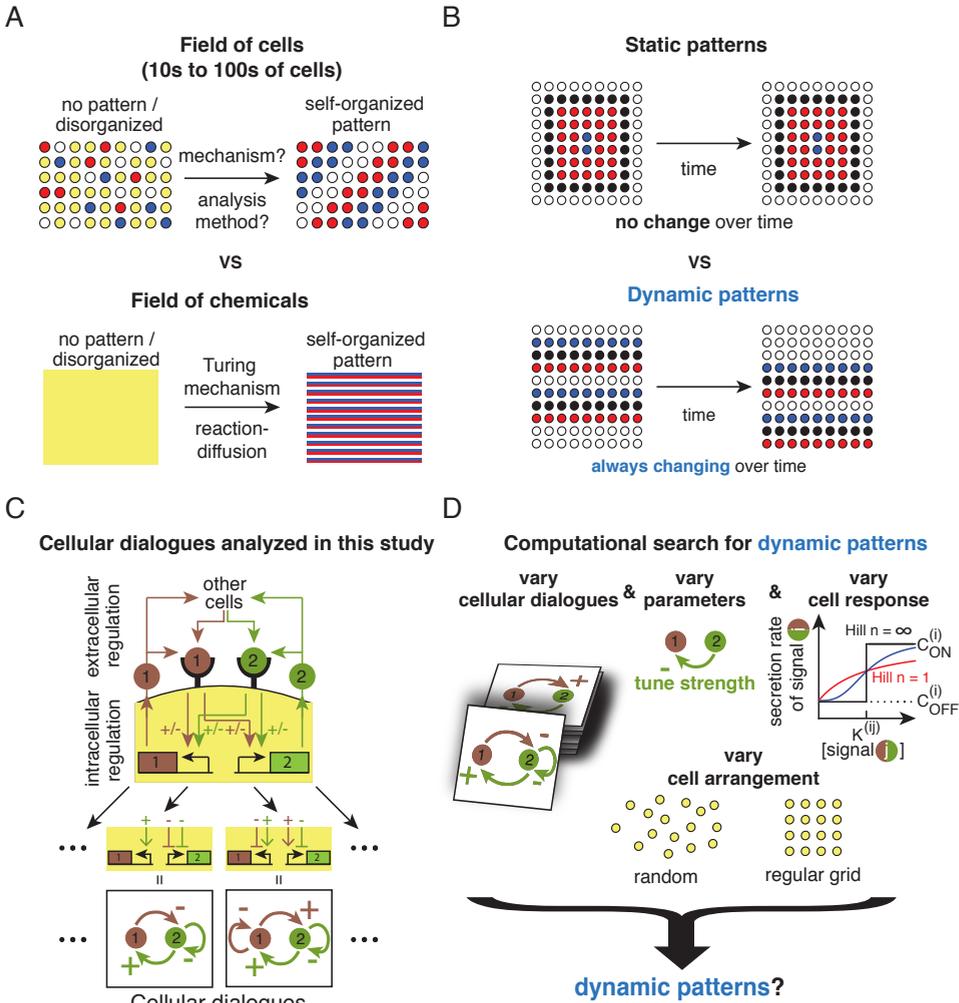
We first considered cells that digitally respond to each molecule (as motivated in Chapter 3.1): a cell secretes “molecule- $i$ ” at either a low rate (“OFF” state for molecule- $i$ ) or a high rate (“ON” state for molecule- $i$ ). If molecule- $j$  activates (represses) molecule- $i$ , then a cell becomes ON (OFF) for molecule- $i$  if and only if it senses a concentration of molecule- $j$  that is *above* a set threshold concentration. The digital cells also have a reporter gene for each molecule, which we call genes “1” and “2”, that are also either ON or OFF to reflect the secretion state of its corresponding molecule (Figure 5.1C – brown and green boxes). In our simulations, we assigned a distinct color to each of the four states, which are (ON for gene-1, ON for gene-2), (ON, OFF), (OFF, ON), and (OFF, OFF).

We began each simulation by randomly assigning the four gene-expression states (i.e., four colors) to each cell so that the gene expression levels were spatially uncorrelated. Thus, the field of cells initially did not exhibit any spatial organization. We quantitatively verified this with a “spatial index” metric which is a weighed spatial autocorrelation function that is zero when cells are completely, spatially disorganized and increases towards one as the cells become more spatially organized (see Equation S5.13 and Figure S5.4). We then observed how each cell’s state (i.e., four colors) changed over time to determine whether a spatial pattern formed and, if so, what type of a pattern formed. For each cellular dialogue, we fixed the values of all parameters (e.g., threshold concentrations, secretion rates for each molecule), and then ran large numbers of simulations with different initial conditions (see Section S5.4.2). We screened a wide range of parameter values for every possible cellular dialogue (see Section S5.4.2). We first performed such a computational search with immobile digital cells that were placed on a regularly spaced lattice. We will first describe these results in the next sections before explaining how these results change when we relax the constraints – by randomly displacing cells so that they no longer form a regular lattice, having each cell continuously move, allowing the Hill coefficient to be any finite value (i.e., analogue instead of digital response), and including gene-expression noise (Figure 5.1D).

### 5.2.2. CELLULAR DIALOGUES ENABLE SELF-ORGANIZATION OF WIDE ARRAY OF DYNAMIC PATTERNS

The computational search revealed a wide variety of dynamic patterns, from never-ending traveling waves (Figure 5.2A) to complex patterns consisting of wavelets that evolved over time in an erratic, complex manner (Figure 5.2B). All patterns self-organized from completely disorganized fields of cells by their ON/OFF-states becoming more spatially correlated over time (Figure 5.2A–B). The time taken to self-organize widely varied and depended on the type of pattern formed. For example, if we assume that a gene-expression change such as an ON-cell becoming an OFF-cell takes one minute – this is one time-step of a simulation and every cell synchronously changes their ON/OFF states – then horizontal waves could take nearly six hours to form (Figure 2A) whereas the constantly changing, complex whirlpool of wavelets would not show any signs of settling into any pattern that cyclically repeats itself even after a week or longer (Figure 5.2B). Since the simulations are deterministic for now – we will later add gene-expression noise – once a simulation reproduces a spatial configuration that it had before, the cell population has formed a dynamic pattern that periodically repeats itself forever.

The dynamic patterns that we uncovered differed in their shape, complexity, and movements (Figures 5.2C–J, and Section S5.5.1). Among these, the most prominent were recti-



**Figure 5.1: Computationally screening cellular dialogues to find ones that enable dynamic patterns to form.** (A) Pattern formation by cells versus chemicals. (Top) Mechanisms by which an initially disordered field of a mesoscopic number of cells (~hundreds to thousands) become more ordered through cell-cell communication (right panel) remain poorly understood, as is the method to analyze this complex self-organization dynamics. (Bottom) A field of chemicals or a continuum of cells (large number of tightly packed cells) initially having no pattern (left) can form a pattern (right) without pre-existing morphogens. This is usually modelled by reaction-diffusion equations and can be understood through the Turing mechanism. (B) Static versus dynamic patterns. (Top) Static patterns do not change over time. (Bottom) In dynamic patterns, a structure changes over time without ever stopping (e.g., shown here is a traveling wave). (Caption continued on next page.)

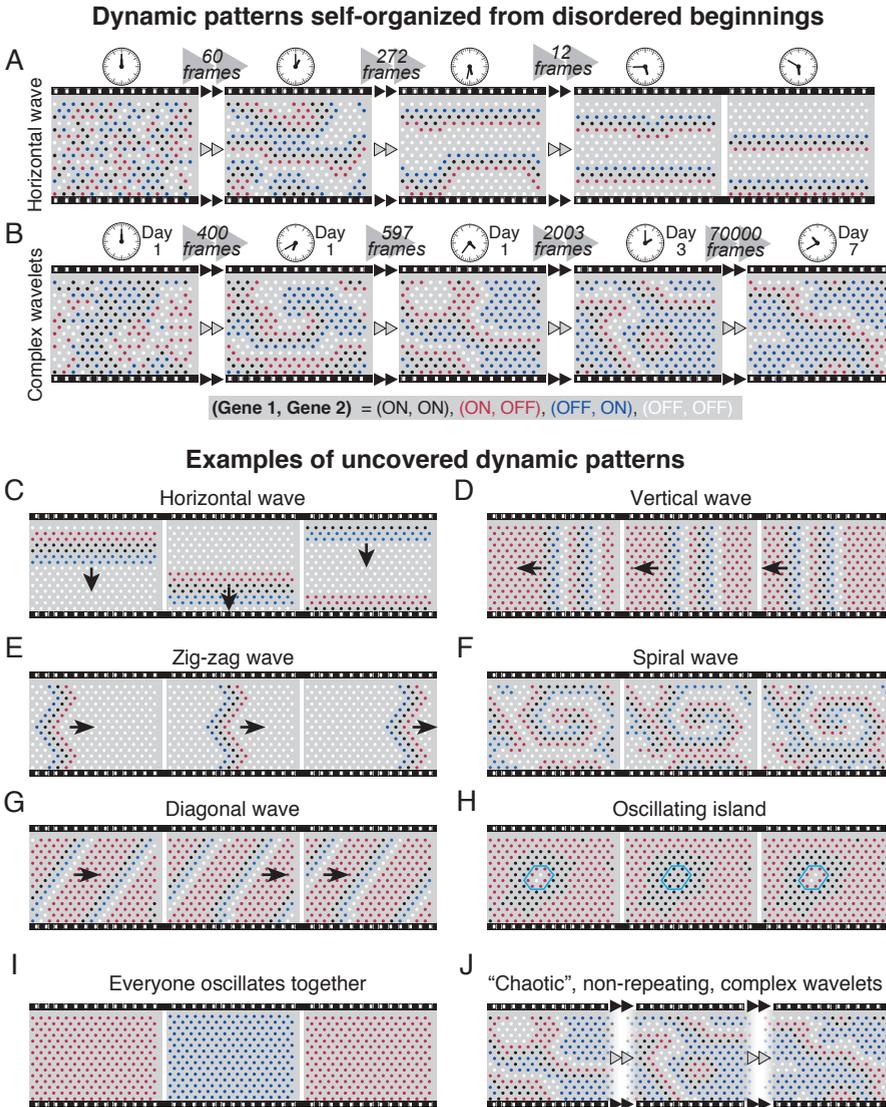
**Figure 5.1 (previous page):** (C) Schematic of cellular dialogues. Brown (molecule 1) and green (molecule 2) circles are ligands that bind to their cognate receptors on the cell membrane. Ligand-bound receptors trigger intracellular signal transductions that either positively or negatively regulate the production and secretion of molecules 1 and 2 (molecule 1 can self-promote or self-repress its own secretion while also regulating the secretion of molecule 2, and vice versa). Bottom row shows graphic representation of cellular dialogues. (D) Elements that we varied in simulations: cellular dialogues of all possible topologies, the values of the parameters for each cellular dialogue, and spatial arrangement of cells. Our study first begins with an infinite Hill coefficient (i.e., digital response to each of the two signaling molecules) and a regular lattice. After reporting the outcomes of these simulations, we report the result of relaxing these two constraints and well as other elements not depicted.

## 5

linear traveling waves and spiral waves, both of which have high degrees of spatial order (Figures 5.2C–F). In the case of traveling waves – which can be oriented horizontally, vertically, or diagonally (Figures 5.2C–D, and 5.2G) and have a straight or bent shape (Figures 5.2D–E) – a rigid shape moves across space over time. Since the simulations were deterministic and the system had periodic boundary conditions, if the simulation revisits an earlier spatial configuration, then it would periodically and forever repeat the same dynamics from then on. In the case of traveling waves, this meant that the waves perpetually propagated, disappearing at one edge of the field and then appearing at the opposite end. This behavior also applies to patterns that do not propagate over space, but rather, oscillate in time. In some cases, such oscillations were limited to a few cells that formed an island (Figure 5.2H) whereas in others, every cell in the field oscillated together (Figure 5.2I). In particular, an island of cells could oscillate in such a way that individual cells oscillated with different periods (Figure 5.2H), causing the entire island, as a collective entity, to display a complex oscillation with a period larger than four timesteps. We call this a “complex” oscillation because the simplest oscillation would involve all cells in the island having the same gene-expression state that oscillates with a period of at most four timesteps, since a cell can have at most four distinct gene-expression states (i.e., (ON/OFF, ON/OFF)). Finally, some cellular dialogues yielded temporally non-repeating, complex patterns consisting of whirlpools of wavelets that evolved over time in an erratic manner (Figure 5.2J) which, in many cases, transiently existed for tens of thousands of timesteps before the cells formed temporally repeating, well-defined dynamic patterns such as horizontal waves.

### 5.2.3. COMMON STRUCTURAL ELEMENTS IN CELLULAR DIALOGUES THAT GENERATE DYNAMIC PATTERNS

The wide array of dynamic patterns that we observed fall into two categories (Figure 5.3A): (1) dynamic *temporal* patterns, in which cells periodically oscillate over time but do not propagate information over space (e.g., Figures 5.2H–I), and (2) dynamic *spatial* patterns, in which cells propagate information over space in the form of a well-defined



**Figure 5.2: Examples of self-organized dynamic patterns found through computational screening.** In all the figures shown here, a cell (drawn as a circle) can have four colors. Each color represents a distinct gene-expression state, (gene 1 = ON/OFF, gene 2 = ON/OFF): Black means (ON, ON), red means (ON, OFF), blue means (OFF, ON), and white means (OFF, OFF). In all the simulations, a field of cells starts with a completely spatial disordered configuration — there is no correlation between neighboring cells' gene-expression states — as exemplified by the leftmost picture shown in (A). (Caption continued on next page.)

**Figure 5.2 (previous page):** (A) Traveling wave of horizontal bands. Snapshots of the formation process shown at different stages of a simulation. Assuming that one timestep in the simulation takes one minute, the clocks show time passed from noon (beginning of the simulation). (B) Complex pool of multiple wavelets formed, starting with a spatially disorganized field of cells. Snapshots at different stages of the simulation are shown. Assuming that one timestep represents one minute, the clock and the days elapsed indicate at which timesteps in the simulation the snapshots are taken. (C–J) Each filmstrip shows three non-contiguous snapshots of a moving, dynamic pattern that formed, starting from a spatially disorganized configuration (not shown, see examples in the first snapshots in A). Where shown, the arrows represent the direction of travel. The dynamic patterns are: (C) a single traveling horizontal band, (D) traveling vertical bands, (E) a traveling zig-zag band, (F) a spiral wave, (G) traveling diagonal bands, (H) a small island of cells (enclosed in the blue hexagon) oscillating over time while all cells outside the island remain static, (I) every cell oscillates between red and blue with period 2, and (J) seemingly chaotic, never-ending dynamics in which multiple wavelets form and meet and annihilate each other, with the pool of wavelets constantly evolving and never repeating the same configuration throughout the simulation.

## 5

shape (e.g., a wave front) that moves from one part of the field to another, often from one edge to the other edge of the field (e.g., Figures 5.2C–F). There are 44 distinct cellular dialogues in total (see Section S5.4.1) that we could group into three categories: (1) those that cannot form any dynamic patterns, (2) those that can form only dynamic temporal patterns, and (3) those that can form both dynamic spatial patterns and dynamic temporal patterns. To categorize them, we developed a method to deduce, for each cellular dialogue, all possible ways that a cell's state (ON/OFF, ON/OFF) can change over time. Concretely, we constructed a directed graph for each cellular dialogue (see Section S5.5.2) which has four nodes — one for each gene-expression state — that are connected by edges with directions that represent the allowed transitions between the nodes. We deduced how some of the directed edges become inaccessible while others become accessible as we change the cellular dialogue's parameter values (Figure S5.7). Then, following the directed edges from node to node yields all possible ways that a cell's gene expression can change over time. By looking for graphs that contained cyclic paths, we identified cellular dialogues and ranges of their parameter values that can potentially sustain dynamic patterns if they were to form. Since self-organization of dynamic patterns can only occur for parameter values that can sustain dynamic patterns in the first place, we only had to check these values in simulations to see if they led to dynamic patterns. This method thus vastly reduced the range of parameter values that we had to screen. For each cellular dialogue, we generated a large set of random parameters and ran many simulations (see S5.4.2), each starting with a different and randomly generated gene-expression pattern. We checked whether each of these simulations yielded a dynamic pattern using automated methods (see Section S5.4.2).

We discovered that cellular dialogues, when grouped into the three categories mentioned above, form distinct tree structures (Figures 5.3B–D) in which a node denotes a particular cellular dialogue and an edge connects two nodes if one node (cellular di-

alogue) comes from the other node (another cellular dialogue) by adding or removing one regulatory interaction. The fact that tree structures emerged, which link the different cellular dialogues together if they form the same type of patterns, suggests that there may be common elements in the cellular dialogues that belong to the same tree. Indeed, we found that all ten cellular dialogues (Figure 5.3B) that can only generate static configurations, and no dynamic patterns at all, consist of two molecules that do not mutually regulate each other and also do not have any self-repressions. We also found that twenty-six cellular dialogues can produce dynamic temporal patterns but not dynamic spatial patterns (Figure 5.3C). Their common feature is that they all contain a self-repression and/or a mutual feedback of the same sign (i.e., both molecules either activate or repress each other's production). The sole exception to this rule, within this family of cellular dialogues, is cellular dialogue 14 (Figure 5.3C). Cellular dialogue 14 consists of an activator-inhibitor pair, whereby one molecule promotes the production of the second molecule, which in turn represses the production of the first molecule. Here, neither molecule regulates its own production. However, all eight cellular dialogues that one can obtain from cellular dialogue 14 by adding one or more self-interactions can yield dynamic spatial patterns, in addition to dynamic temporal patterns (Figure 5.3D). We could further divide these eight cellular dialogues into two classes: ones that contain only self-repressions (Figure 5.3D – blue boxes) and ones that contain at least one self-activation (Figure 5.3D – red boxes). The three cellular dialogues that contain only self-repressions produce dynamic spatial patterns in which the moving shape periodically changes its gene-expression composition (Figure S5.6). In contrast, the five cellular dialogues that contain at least one self-activation yield dynamic spatial patterns such as traveling waves (Figures 5.2C–G) in which the pattern moves across the field of cells without changing in shape or composition.

#### GROUPING CELLULAR DIALOGUES BASED ON HOW FAST THEY FORM PATTERNS IS EQUIVALENT TO GROUPING THEM BASED ON THEIR SHARED STRUCTURAL ELEMENTS

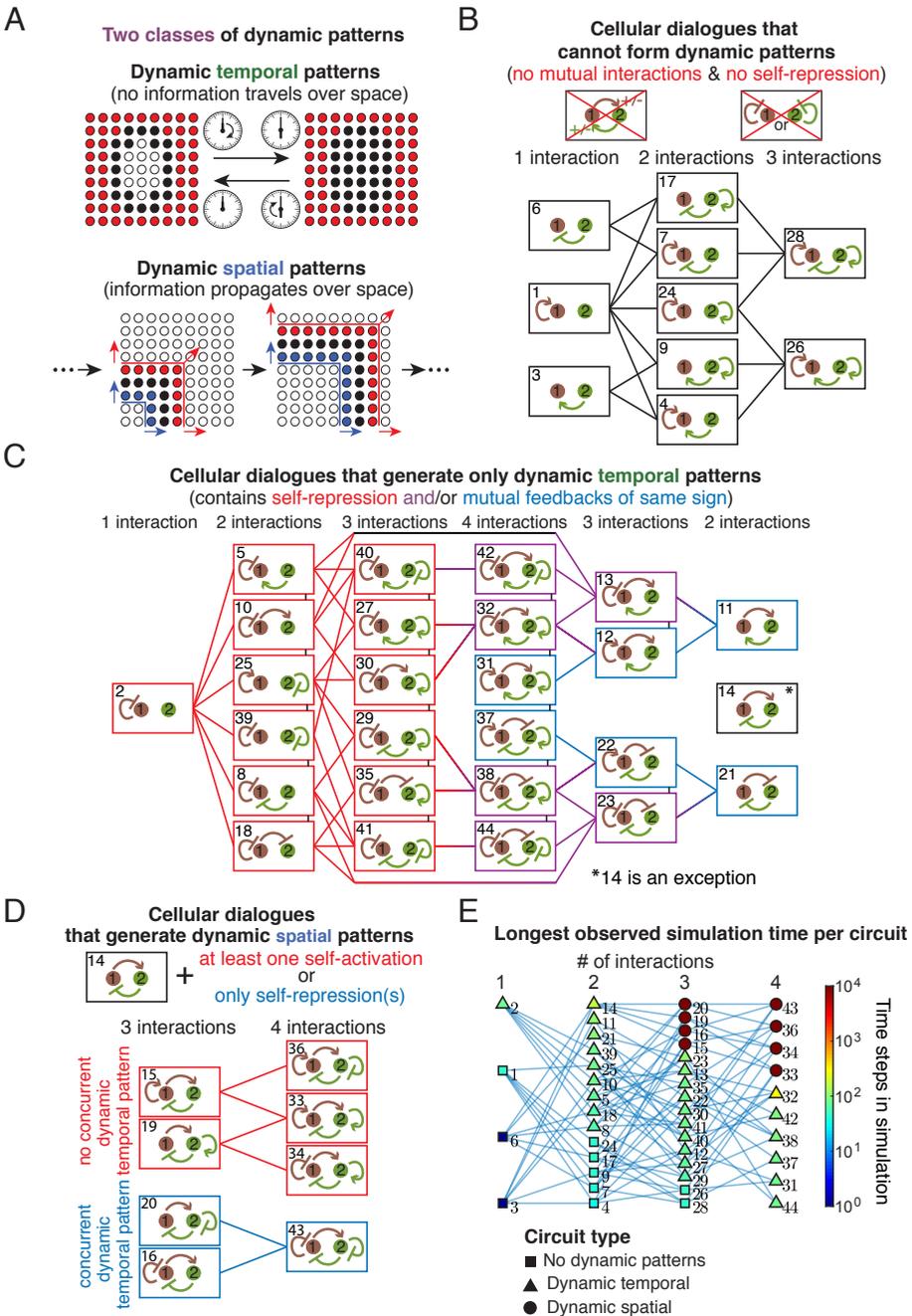
We discovered that if we analyze the typical times or the longest time that a cellular dialogue takes to form a pattern (static configuration or a dynamic pattern), and then group the cellular dialogues based on those times, then we would identify the same three categories of cellular dialogues (Figure 5.3E and Figure S5.1). Specifically, all eight cellular dialogues that can form dynamic spatial patterns stood out as taking the longest times to form patterns compared to the other cellular dialogues, by at least about 100-fold longer durations (Figure 5.3E – circles). As we will later discuss, we found that these long self-organization times (~1 week if one time-step represents one minute) are due to complex dynamics that is intrinsic to the pattern-formation process. We found that all cellular dialogues that cannot form dynamic spatial patterns but do form dynamic

temporal patterns take less times to form patterns, by at least a 100-fold less, than the ones that form dynamic spatial patterns (Figure 5.3E – triangles). Finally, we discovered that the cellular dialogues that cannot form any dynamic patterns and thus only form static configurations – some of which are highly organized patterns – require the least amounts of time to form these configurations (Figure 5.3E – squares).

#### 5.2.4. ANALYTIC FRAMEWORK EXPLAINS HOW CELLS COLLECTIVELY SUSTAIN DYNAMIC SPATIAL PATTERNS

To explain why certain cellular dialogues enable cells to sustain the dynamic spatial patterns after having formed them, we developed a theory that does not use simulations and still correctly predicts *when* dynamic spatial patterns occur and explains *how* the cells sustain them (Figures 5.4A–C). The key idea behind this analytical approach is that many dynamic spatial patterns, from the complex whirlpools of wavelets to spiral waves, share a common structure: one can build diverse dynamic spatial patterns by gluing together multiple rectilinear waves (i.e., horizontal, vertical, and bent waves). Thus, if we can understand how cells can sustain rectilinear waves, we can piece them together to understand the more complex shapes that are built out of them. Each rectilinear wave has six distinct layers of gene-expression states (Figure 5.4A). Three of the layers – “front”, “middle”, and “back” (Figure 5.4A – red, black, blue cells) – constitute the wave itself and continuously move forward while the other three layers – “exterior front”, “exterior” and “exterior back” – consist of all the other cells. After one timestep, each layer adopts the identity of the layer just behind it (e.g., the exterior-front layer, which is just in front of the front layer, becomes the front layer) (Figure 5.4C). This must occur at every timestep in order for the wave to continuously propagate, meaning that the concentrations of the two molecules within each layer must coordinately change so that the layers can synchronously move forward. We developed a method to estimate the concentrations of the molecules in each layer (Figure 5.4B and Section S5.5.3).

Using the analytical approach, we derived six mathematical inequalities, one for each layer, that must all be satisfied in order for the concentrations of the two molecules to coordinately change to enable the rectilinear wave to propagate (Figure 5.4C and Section S5.5.3). The inequalities impose relationships among the different parameters of the cellular dialogues, such as the maximal secretion rates and sensing thresholds (Figure 5.1D). By solving these inequalities, we found that only five cellular dialogues – the exact same ones that we computationally identified – can satisfy all six inequalities and thus generate non-oscillatory dynamic spatial patterns (i.e., the ones that do not involve concurrent dynamic temporal patterns) (Figure 5.3D – red boxes). In accordance with the computational screening, the analytical approach revealed that only two types of



**Figure 5.3: Computational search revealed tree structures that group cellular dialogues based on their ability to generate either static patterns, dynamic temporal patterns, or dynamic spatial patterns.** (A) Two classes of dynamic patterns. (Top): Dynamic temporal patterns repeat themselves over time without transmitting information across space. (Bottom): Dynamic spatial patterns involve cells that transmit information over space through a coherent structure that moves across the field. (Caption continued on next page.)

**Figure 5.3 (previous page): (B–D)** Tree diagrams show a full classification of all 44 unique, non-trivial cellular dialogues into three distinct classes. In each tree diagram, a cellular dialogue is a leaf (box) that is joined by branches to other cellular dialogues. As one moves from one leaf to the next, an edge is either removed or added to the cellular dialogue. (B) Tree diagram showing all cellular dialogues that cannot generate any dynamic patterns. All cellular dialogues here lack mutual interactions and self-repressions. (C) Tree diagram showing all cellular dialogues that can generate dynamic temporal patterns but not dynamic spatial patterns. These all have either a self-repression (red boxes), a mutual interaction of the same sign (blue boxes), or both (purple boxes). Cellular dialogue 14 is an exception — it has mutual interactions of different signs and no self-interactions. (D) Tree diagram showing all cellular dialogues that can generate dynamic spatial patterns — these can all also generate dynamic temporal patterns. These are all generated by adding at least one additional self-interaction to cellular dialogue 14. Cellular dialogues in the five red boxes have at least one positive feedback loop, and can generate non-oscillatory dynamic spatial patterns (e.g., traveling waves). Cellular dialogues in the blue boxes have only negative self-interactions and produce dynamic spatial patterns but always with a concurrent dynamic temporal pattern (e.g. a traveling wave where the cells oscillate simultaneously) (see Figure S5.6 for examples). (E) The maximum observed simulation time is a metric that naturally separates the three classes of cellular dialogues (B–D) (see Figure S5.1 for other metrics). A node represents a cellular dialogue and the node's shape represents the type of cellular dialogue (one of the three (B–D)). A node's color indicates the longest observed simulation time among a large set of simulations that were performed with different parameters.

## 5

rectilinear waves are possible, each differing by which gene-expression state is assigned to each layer: all cellular dialogues with cellular dialogue 15 as the common motif (i.e., molecule-1 promotes its own secretion) generate one type of rectilinear wave (Figure 5.4D – top row) while the others, having cellular dialogue 19 as the common motif (i.e., molecule-2 promotes its own secretion), generate the other type of rectilinear wave (Figure 5.4D – bottom row). As an exception, cellular dialogue 33 can generate both types of traveling waves because nested in it are both cellular dialogues 15 and 19 as subgraphs.

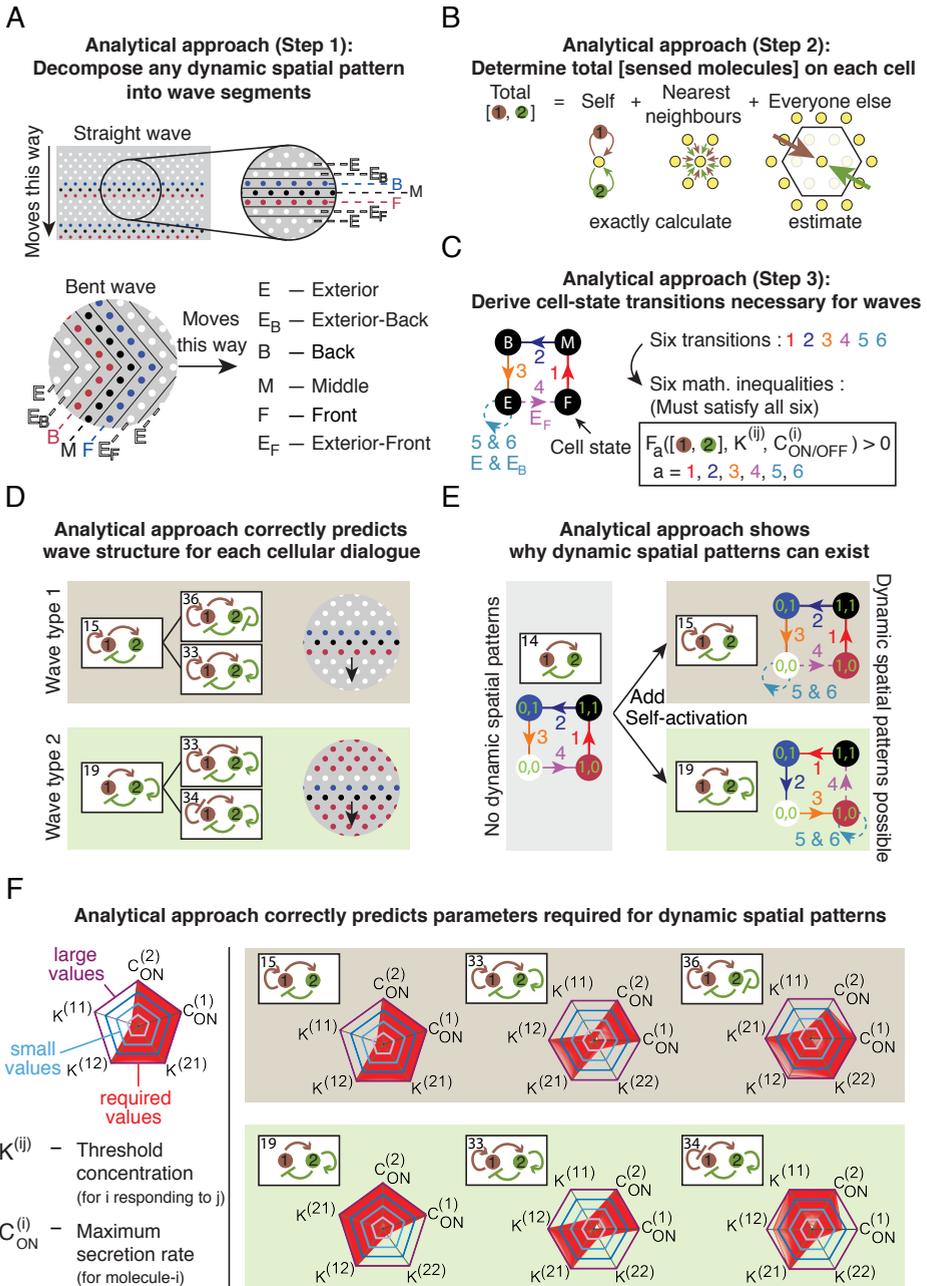
To understand why only these five cellular dialogues (Figure 5.4D) can generate dynamic spatial patterns, we considered the directed-graph representation of the cellular dynamics that we introduced earlier (Section S5.5.2). For a wave, the directed graph must contain a cyclic path that goes through all four nodes – one node for each gene-expression state – since an exterior cell must eventually become a front-layer cell, then a middle-layer cell, then a back-layer cell, and then finally an exterior cell again (Figure 5.4C). Cellular dialogue 14, which is the backbone of all five cellular dialogues that generate dynamic spatial patterns (Figure 5.3D – red boxes), can potentially produce a cyclic graph with these four nodes (Figure 5.4E – left panel) as long as they permit parameter values that allow each cell to cyclically traverse through the nodes. This is because starting with a gene-expression state of  $(1, 0)$  – where the 1 means ON-state for molecule-1 and the 0 means OFF-state for molecule-2 – may lead to  $(1, 1)$  due to molecule-1 promoting molecule-2 secretion, which then may lead to  $(0, 1)$  due to molecule-2 repressing molecule-1 secretion, which then may lead to  $(0, 0)$  due to there being not enough molecule-1 for promoting molecule-2 secretion, and finally, this may lead back to the

starting state,  $(1, 0)$ , due to there being not enough molecule-2 for inhibiting molecule-1 secretion. However, such a cycle through the four nodes alone is insufficient for sustaining a wave because the exterior cells must remain as exterior cells unless they are adjacent to the front or back layer (Figure 5.4C). But if the exterior cells have state  $(0, 0)$  and the front-layer cells have state  $(1, 0)$ , then the exterior cells near the front layer (i.e., the exterior-front cells) would sense more molecule-1 than the exterior cells that are further away from the wave. Modifying cellular dialogue 14 by having molecule-1 promoting its own secretion, as in cellular dialogue 15, would create the possibility of the exterior-front cells activating molecule-1 secretion and thus transition to  $(1,0)$  at the next timestep, thereby becoming a front-layer whereas the exterior-layer cells remain in the  $(0, 0)$  state (Figure 5.4E – top right). A similar reasoning also yields an analogous result for cellular dialogue 19 (Figure 5.4E – bottom right).

To realize the qualitative scenario described above, a cellular dialogue must contain parameter values that satisfy all six inequalities that we derived (Figure 5.4C). We found that the five cellular dialogues indeed admit such parameter values and that these values — obtained through the analytical approach — nearly perfectly match those found in the computational screen (Figures S5.9–S5.10). We can represent these parameter values as spider charts (Figure 5.4F), which show that each of the five cellular dialogues can realize dynamic spatial patterns with parameter values that vary over many orders of magnitude. The spider charts also geometrically reveal a common feature among the five cellular dialogues: the threshold concentration must be low for a molecule that promotes its own secretion (Figure 5.4F – note the inward indentations in the red spider webs along the axes that represent the threshold concentrations). This makes sense because, for all types of rectilinear waves (Figure 5.4D), the exterior-front cells need to turn on the secretion of a molecule that promotes its own secretion by sensing it from the other layers and having a low activation threshold for that molecule would facilitate this. Taken together, our analytical approach unveiled how cells can sustain dynamic spatial patterns.

### 5.2.5. SELF-ORGANIZATION OCCURS THROUGH A THREE-STAGE, “ORDER-FLUCTUATE-SETTLE” MECHANISM

We now turn to the self-organization process itself. Given that many of the dynamic spatial patterns are traveling waves and that more complex dynamic spatial patterns can be built from gluing together multiple rectilinear waves, we focused on traveling waves and the core features of their self-organization process. Our simulations revealed that traveling waves form in three stages (Figure 5.5A). First, a field of cells whose gene-expression levels form a completely disorganized spatial configuration rapidly becomes more spa-



**Figure 5.4: Analytic framework predicts and explains how cells can sustain dynamic spatial patterns.** (A–C) Three-step overview of an analytic (pen-and-paper) approach to understanding the simulations (see Section S5.5.3). (A) Step 1: Decompose straight (top) and bent (bottom) waves into distinct layers of cells. Cells of the same layer have the same gene-expression state. (Caption continued on next page.)

**Figure 5.4 (previous page):** (B) Step 2: Estimate the total concentrations of molecules that a cell senses by exactly calculating the portions of those concentrations that are due to the cell itself and its nearest neighbors, and by approximating the portions of the total concentrations that are due to further-away cells, and by approximating the portions of the total concentrations that are due to further-away cells. (C) Step 3: (right) Directed graph-representation showing how a cell must transition to distinct layers shown in (A) at each timestep, which is explained by six mathematical inequalities that are derived through step 2 (see Section S5.5.3). (D) Numerically solving the six inequalities in (C) shows that only two types of waves, shown here, are possible and which cellular dialogues can produce them (cellular dialogues 15, 36, and 33 for wave type 1; cellular dialogues 19, 33, and 34 for wave type 2). (E) Adding self-activation to cellular dialogue 14 yields, in the left column, cellular dialogues 15 and 19. Directed graph-representation showing the gene-expression transition of a cell for each cellular dialogue is shown (see Section S5.5.2). (F) Parameter values that allow for sustaining of rectilinear waves, when represented as red points, form a dense region (red region) as shown in these spider charts. These parameter values satisfy the six inequalities derived by the analytic theory (C) (see Figure S5.10C for a direct comparison with parameter values found purely through computational search). The spider charts show the following parameters: threshold concentrations  $K^{(i,j)}$  for each molecular interaction and the maximum secretion rate  $C_{ON}^{(j)}$  for each of the two molecules.

tially ordered, meaning that the gene-expression levels of neighboring cells tend to become more correlated over time. To quantify the degree of spatial organization, we used a “spatial index” — a metric from our previous work whose value is zero for a completely disorganized spatial configuration and increases towards one as the spatial configuration becomes more organized (see Equation S5.13 and Figure 5.5B – left panel’s inset) [Maire & Youk, 2015a; Olimpio et al., 2018].

In the following discussion, we consider one timestep to represent one minute and express the time in minutes or hours. Then this rapid spatial ordering typically takes less than an hour (Figure 5.5A – green arrow and Figure 5.5B – left panel). At the end of this process, the cells have formed multiple whirlpools of wavelets (Figure 5.5A – frame at 0.33 hours). Thus begins the second stage of self-organization: long-lived complex dynamics — lasting for days or weeks — in which multiple wavelets travel through the field of cells, meeting and annihilating each other, all the while as the cells form new wavelets to replace the destroyed ones (Figure 5.5A – filmstrip from 0.33 hours to 55 hours). During this days-long dynamics, the spatial organization neither stably increases nor decreases — the spatial index erratically (unpredictably) fluctuates over time (Figure 5.5B – left panel; Figure S5.3), which we can see by plotting the Fano factor for the spatial index over time (Figure 5.5C – left panel and Section S5.4.1). The spatial index erratically fluctuating represents multiple wavelets forming and annihilating at various, seemingly random locations and wavelets unpredictably morphing over time, all despite the fact that the simulations are completely deterministic. Crucially, we verified that the same spatial configuration never repeats itself throughout the days-long dynamics which could, in fact, last for weeks or longer if we do not terminate the simulations (i.e., some fields of cells never reach a steady-state and never attain a dynamic pattern within the allotted

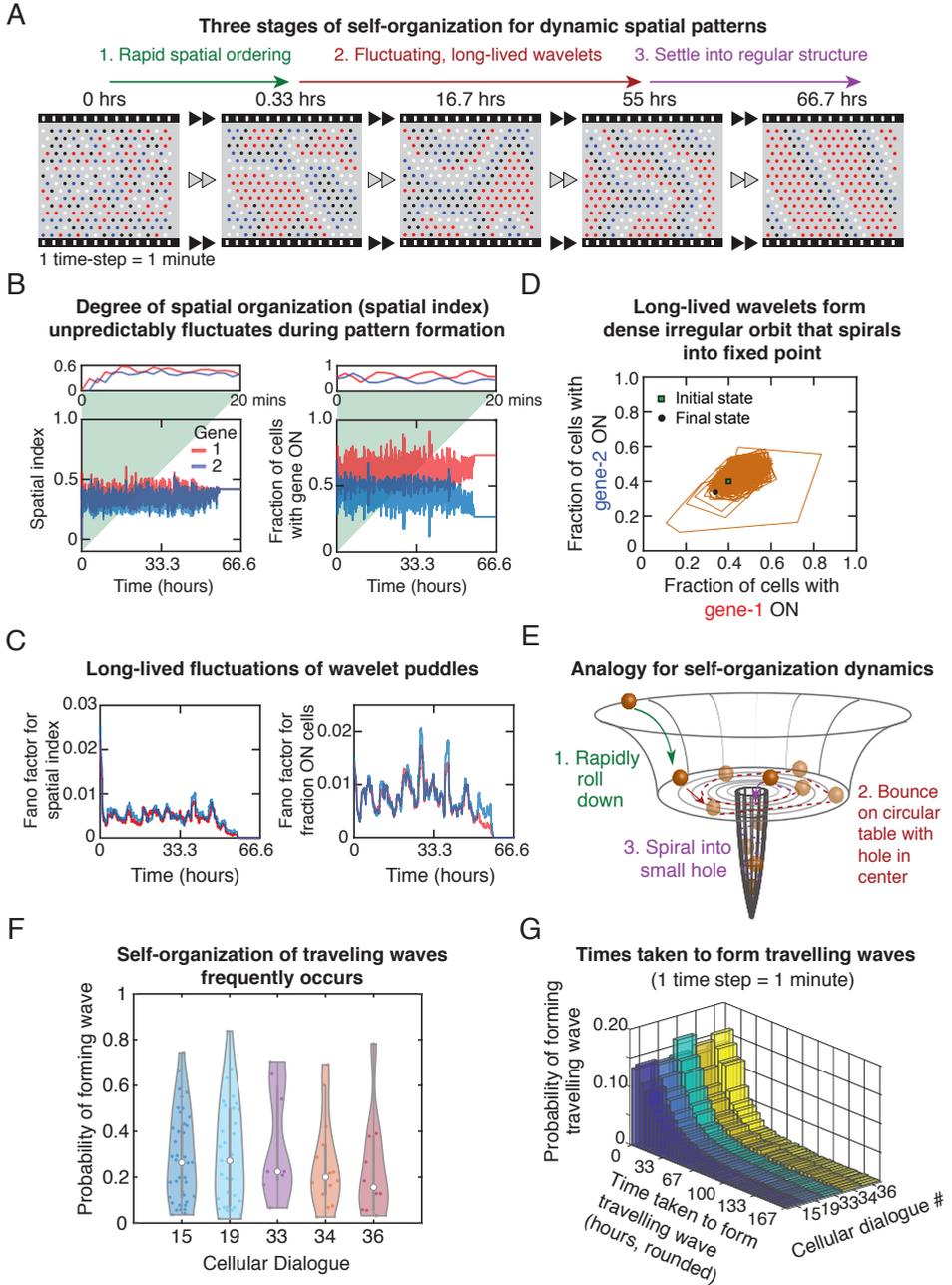
time for the simulations). Such erratic, complex dynamics is followed by the third and final stage of the self-organization process: the wavelets die down and as this occurs, a more rigid, spatially ordered structure that travels as a wave emerges (Figure 5.5A – last frame). During this final process, the spatial index's fluctuations rapidly decay, typically over a few hours. The system then settles into a regular dynamic pattern that repeats itself over time. This is marked by the sudden disappearance of the fluctuations in the spatial index (Figure 5.5B – left figure). This settling process takes a few minutes to several hours (Figure 5.5A – purple arrow and Figure 5.5B – left panel). Leading up to this last stage, there are no clear indications that a well-organized regular shape will emerge. This highlights the erratic, complex nature of the self-organization dynamics.

The spatial index, one for each gene, represents a macrostate variable — a single number that measures how much spatial correlation there is in the expression of a particular gene (see Section S5.4.1). Another macrostate variable is the fraction of cells that have the same gene-expression level (i.e., fractions of cells that have gene-*i* in the ON-state). There are two such fractions, one for each gene. During the self-organization process, these two fractions erratically fluctuate over time — just like the spatial indices — as the wavelets constantly and erratically change their shapes while meeting and annihilating each other for days. Afterwards, the two fractions' fluctuations quickly decay over time — the decay takes a few hours whereas the whole self-organization process takes days — and eventually settle at steady-state values (Figure 5.5B–C: right panel and Figure S5.3). When we view the temporal change of these two fractions as a trajectory in a plane — a phase space — defined by the two fractions, we see an irregular orbit that eventually stops at a single point (Figure 5.5D – black circle). Specifically, a point in the two-dimensional phase space — representing the values of the two fractions at a given time — erratically moves within a restricted region of the plane. If we follow the trajectory with a pencil, we would obtain a jagged curve that densely and nearly entirely fills the whole space within the restricted region that encloses the single point where the trajectory terminates.

The phase-space trajectory described above suggests the following analogy for the self-organization dynamics (Figure 5.5E): a ball quickly rolls down a steep side of a large bowl, speeding up as it does so, until it reaches the bowl's flat bottom. This is the first stage of self-organization in which the decreasing height represents more spatial ordering (Figure 5.5E – green arrow). After reaching the frictionless, flat circular bottom, the ball rapidly bounces off the side walls, like a billiard ball, without ever losing its speed (Figure 5.5E – brown dashed lines). This bouncing ball, which would produce seemingly erratic yet deterministic motion — as Newton's laws of motion are deterministic — represents the second stage of self-organization in which multiple whirlpools of wavelets

are unpredictably created and destroyed. Eventually, the ball finds the small hole, falls into it, and then spirals its way downwards along the side walls of the trench through the hole until it reaches the bottom of the trench (Figure 5.5E – purple arrow). This would represent the third and the final stage of the self-organization. The shape of the bowl and the location of the trench would be determined by the parameters of the cellular dialogue.

In each of the five cellular dialogues that can yield dynamic spatial patterns, we found that, for parameter values that enable dynamic pattern formations, approximately 30% of the initially disorganized spatial-configurations successfully self-organized traveling waves (Figure 5.5F). Moreover, our simulations and the analytical approach revealed that cells can have arbitrarily high parameter values and still form traveling waves, as long as the secretion rates and threshold concentrations are appropriately tuned (Figures S5.10 and S5.12). Our analytical framework presents an optimization strategy for ensuring that cells form traveling waves for the largest possible set of parameter values (Figure S5.11). This strategy depends on balancing how much a cell communicates with itself by capturing back the molecules that it had just secreted (self-communication) with how much a cell communicates with the other cells by sending its secreted molecules to them (neighbor-communication). In short, we found that when the cells are sparsely packed, there is not enough neighbor-communication for sustaining traveling waves. On the other hand, when the cells are densely packed, then cells cannot cycle through a set of gene-expression states — a requirement for dynamic patterns such as waves — because the signaling molecules quickly reach saturating concentrations rather than undergoing the necessary cycles of decreases and increases. This leaves us with intermediate, “goldilocks” density of cells as being ideal for forming and sustaining waves and dynamic spatial patterns (Figure S5.11B). Furthermore, for all five cellular dialogues, we discovered that the probability of forming a traveling wave at a given time is well described by an exponential distribution (Figure 5.5G and Figure S5.8A), with a characteristic decay time of thousands of timesteps (i.e., tens of hours if one timestep is one minute). This strongly suggests that traveling wave formation is a memoryless process whereby at each timestep, the probability that the next timestep yields a traveling wave remains the same regardless of at which timestep the simulation is at. This reflects the fact that watching the simulations that yield a dynamic spatial pattern does not give the observer a sense that the cells are getting anywhere closer to forming a dynamic spatial pattern as time passes (Figures S5.8B–D).



**Figure 5.5: Three-step, “order-fluctuate-settle” process leads to formation of dynamic spatial patterns. (A)** Snapshots of a simulation showing the three stages of a traveling-wave formation – the three stages are described above the filmstrip. Assuming that one time-step of a simulation represents one minute, indicated above each snapshot is the elapsed time in hours. Color scheme for cells is the same as in Figure 5.2. (Caption continued on next page.)

**Figure 5.5 (previous page):** (B) Two macroscopic parameters — the spatial index and the fractions of cells with a particular gene ON — plotted as a function of time for the wave-forming simulation shown in (A). One minute represents one timestep. (Left panel) The spatial index — with magnitude between zero and one — measures the degree of spatial organization (zero means complete disorder, i.e. no spatial correlation in gene-expression among cells, and increasing values correspond to more spatial organization). Inset shows the spatial index rapidly increasing for the first twenty timesteps. Spatial index for gene 1 (red) and gene 2 (blue). (Right panel) Fractions of cells with gene 1 ON (red) and of gene 2 ON (blue) for a typical wave-formation process. Inset shows the first twenty timesteps. (C) For data in (B) and genes 1 (red) and 2 (blue), we used a moving window to compute the moving coefficient of variations in the spatial index (left panel) and in the fractions of cells with the specified gene ON (see Section S5.4.1). (D) For a typical simulation that self-organizes into a traveling wave, we plot the trajectory in phase space formed by the fractions of cells with gene 1 ON and gene 2 ON. The trajectory begins at the square (first timestep of the simulation) and terminates at the circle (last timestep of the simulation). (E) Analogy for the three-stage self-organization process — a billiard ball rolls down a bowl, bounces around on the flat circular bottom, and then fall through a tunnel after finding a small hole drilled into the circular bottom. (F) Probability of forming a traveling wave for each of the five cellular dialogues (detailed results in Figure S5.12). Violin plots showing the non-parametric kernel density (colored distributions), together with the median (white circle), interquartile range (thick vertical line) and 1.5x interquartile range (thin vertical line). Results are obtained by running 500 simulations for each of the parameter sets for which at least one traveling wave formed in the computational screening (see Section S5.4.2). Individual dots represent probabilities for individual parameter sets. (G) Distributions of the time taken to form a traveling waves for each of the five cellular dialogues that enable cells to form dynamic spatial patterns (detailed results in Figure S5.8).

### 5.2.6. DYNAMIC PATTERNS WITH MORE COMPLEX ELEMENTS

We next extended our investigation by relaxing the two main constraints in the simulations — having an infinite Hill coefficient and cells on a regular lattice. We modified the simulations by separately adding four elements (Figure 5.6A and Section S5.4.3): (1) stochastic response to the signaling molecules (Figure 5.6A – top left), (2) a sigmoidal response function characterized by a finite Hill coefficient (i.e., cells no longer digitally respond to the signaling molecules) (Figure 5.6A – top right), (3) randomized locations of cells instead of each cell residing on a regular lattice (Figure 5.6A – bottom left), and (4) random (diffusive) motion of each cell (Figure 5.6A – bottom right). We tuned each element and asked two questions: (1) Can the cells still form traveling waves if they start with a completely disordered spatial configuration? (Figure 5.6B – top) — this probes the self-organization capability — and (2) can the cells still sustain traveling waves after forming them? (Figure 5.6B – bottom) — this investigates whether dynamic spatial patterns can be sustained once formed. In general, we found that cells could still form a wide range of dynamic spatial patterns with the four additional elements (Figure 5.6C). For example, we discovered that cells under the influence of a moderate noise could form a band that travels as a wave despite a number of cells stochastically obtaining the “wrong” (incoherent) gene-expression state. In this case, the wave thus propagates while stochastically evolving (Figure 5.6C – top left). As another example, we discovered that even when we randomly arrange cells in space, instead of on a regular lattice, the cells could still form never-ending, complex wavelets (Figure 5.6C – bottom left).

By running many simulations for each of the four complex elements, we discovered that the dynamic spatial patterns that we previously observed, on a regular lattice with an infinite Hill coefficient (Figure 5.2), still formed as long as the amount of the deviation introduced by the four elements, relative to the regularity of the lattice and the infinite Hill coefficient, was non-negligible but not too large (Figure 5.6D). For instance, we found that, with a moderate noise, dynamic spatial patterns continued to form and persist (Figure 5.6D–E – top left and Figure S5.5). The probability that an initially disordered configuration morphed into a traveling wave became higher with moderate noise, compared to not having any or low noise, indicating that noise can drive the system towards more ordered states — a phenomenon also observed for static patterns in an earlier work [Olimpio et al., 2018]. To account for this observation, we extended our theory, which we developed for explaining wave propagation without noise (Figures 4A–C), to now include noisy gene-expression. Using this extended theory, we calculated the probability that a wave, after forming, “survives” for a given amount of time. This probability closely matched the actual fraction of simulations in which waves survived (Figure S5.13 and Section S5.5.4).

## 5

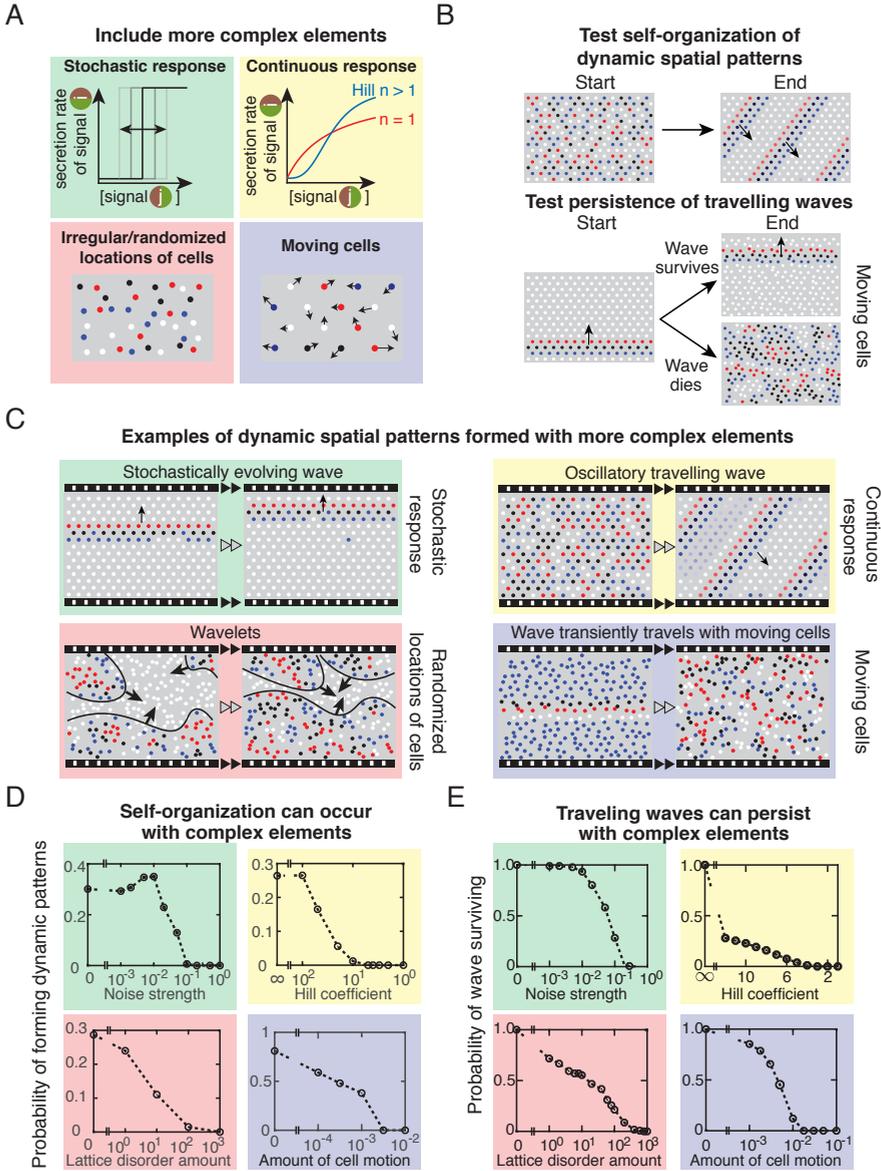
By varying the Hill coefficient over a wide range, we discovered that dynamic patterns can form for finite Hill coefficients of values  $\sim 4$  or higher (Figure 5.6D – top right and Figure S5.5). However, these did not typically include “pure” traveling waves that neatly decompose into the previously identified layers. Moreover, an already-formed traveling wave — as in the case of a simulation that starts with a wave — could persist for Hill coefficients of values down to  $\sim 3$  (Figure 5.6E – top right). These results indicate that a finite Hill coefficient is mainly detrimental to the self-organization of traveling waves whereas it is less detrimental to the cells’ ability to sustain a traveling wave once it is formed. With a Monte Carlo algorithm that randomly displaces the cells and quantifies the amount of resulting “lattice disorder” (see Section S5.4.3), we found that dynamic spatial patterns still formed and persisted even with a high degree of spatial disorder (Figures 5.6D–E – bottom left and Figure S5.10). Even with saturating amounts of spatial disorder, we still observed self-organized wavelets that propagated, albeit with a lesser degree of regularity than in a regular lattice (Figure 5.6C – bottom left). When we allowed the cells to diffusively move — we tuned the cells’ motility by adjusting the diffusivity of their Brownian motion (see Section S5.4.3) — we found that large-scale, uncoordinated motion of the cells prevented any kind of dynamic spatial patterns from stably propagating, as large variations between the local environments of individual cells tended to diminish the cells’ ability to spatially propagate information (Figure 5.6D – bottom right). However, we found that motile cells could still propagate waves, once formed, for an extended amount of time before the wave disintegrated even when the cells had

a high degree of diffusive motion (Figure 5.6E – bottom right). Together, these results strongly suggest that diffusively moving cells can sustain traveling waves as long as the waves travel sufficiently rapidly (i.e., compared to the cells' average speed).

We also studied two more complex elements. First, we considered the influence of a spatial gradient of parameter values on traveling-wave formation (see Section S5.4.3). Researchers have suggested that spatial gradients of parameter values can influence the orientation of Turing patterns such as stripes [Hiscock & Megason, 2015]. Similarly, we observed that a spatially varying parameter, having a simple step-function profile over space, can influence the direction in which the traveling waves moved after forming: the waves tended to align perpendicularly to the gradient (Figure S5.5). Second, whereas until now the cells integrated the two signals with an AND-logic scheme — both molecules were required for activating or repressing gene expression — we repeated the computational search (Figure 5.1D) but now with an OR-logic scheme in which only one of the molecules is required for activation or repression of a gene (see Section S5.4.1). We found that the OR-logic scheme yields exactly the same groupings of cellular dialogues as in the AND-logic scheme in terms of the three classes of patterns that they generate — static, dynamic temporal, and dynamic spatial patterns (Figure S5.2). But we discovered that the OR-logic scheme produces a different “wave structure” (Figure 5.4D) than the AND-logic scheme (Figure S5.2).

### 5.3. DISCUSSION

The dynamic-pattern forming cellular dialogues that we identified include some that have been experimentally observed to yield patterns. They all have interlocked positive and negative feedbacks (Figure 5.3D). Researchers have found that, without any cell-cell communication, such interlocked feedbacks can cause gene-expression levels to robustly oscillate temporally [Stricker et al., 2008, Tsai et al., 2008, Li et al., 2017]. Researchers have also synthetically engineered a quorum-sensing gene circuit resembling cellular dialogue 20 (Figure 5.3D) and observed that the cells' gene-expression levels synchronously oscillate over time and, under certain conditions, spontaneously form traveling waves [Danino et al., 2010]. More generally, the activator-inhibitor structure of cellular dialogue 15 is qualitatively similar to the structure of the FitzHugh-Nagumo (FHN) model, which describes excitable systems such as cells whose biomolecule concentrations oscillate over time and/or form traveling waves [Gelens et al., 2014; Sgro et al., 2015; Hubaud et al., 2017]. Cellular dialogue 15 has an activating molecule that promotes its own production and an indirect negative feedback through the second molecule. This indirect negative feedback is analogous to the slow repression in the FHN model. Similarly, the interlocked positive-negative feedback loops of the dynamic-pattern forming



**Figure 5.6 (previous page): (D)** Fraction of simulations that form a dynamic pattern as a function of the deviation from the more idealized setting — cells placed on a regular lattice and responding digitally with an infinite Hill coefficient — in which the results for Figures 5.1-5.5 were reported. Four colored boxes, with each color corresponding to colored box in (A) that shows the modified element in the simulations. For each data point, we ran a large set of simulations with a fixed set of initial conditions as we varied the parameter controlling the deviation from our original model, and classified their final states (see Figure S5.5 for a details on finite Hill coefficient and noise). All results here are for cellular dialogue 15. **(E)** Fraction of simulations with cellular dialogue 15 that can sustain a traveling wave for at least one full period after starting with a traveling wave. We took parameter values for which the simulations with simpler elements (i.e., infinite Hill coefficient and cells on a regular lattice) can propagate traveling waves.

cellular dialogues resemble the activator-inhibitor systems that generate Turing patterns [Kondo & Miura, 2010] and resemble the two-gene networks that can generate Turing patterns [Scholes et al., 2019]. But the cells in our simulations do not generate Turing patterns such as stripes or spots of fixed sizes, likely due the large separation of timescales between molecular and gene-expression dynamics in our simulations.

Here we focused on cellular dialogues with two molecules and the two genes that they control. But our software can easily be modified to include multiple — more than two — extracellular molecules and genes as well as arbitrary regulations of those genes (as showcased by our inclusion of finite Hill coefficients). Such extensions would allow one to explore more complex ways that cellular dialogues can mediate dynamic-pattern formations. These extensions, our analytical method for analyzing the simulations, and our results on two-molecule cellular dialogues may provide insights on poorly understood systems in which multiple signaling molecules interact with each other. For many biological systems, the regulatory links among the various molecular players remain unknown (Figure 5.7). For example, researchers have found that three signaling molecules — Fgf, Notch and Wnt — regulate one another during somite formations. But how Wnt and Notch regulate each other so that their levels coordinately oscillate over time remains unknown (Figure 5.7A) [Oates et al., 2012; Harima & Kageyama, 2013; Sonnen et al., 2017]. One may address this question by modifying our software to include three-molecule cellular dialogues and then applying our analysis method to analyze those simulations. Doing so may also help in identifying, in stem cells, the as-yet unknown regulatory links among Bmp, Wnt, and Nodal that lead to self-organized spatiotemporal waves (Figure 7B) [Chhabra et al., 2019]. In the Arabidopsis Thaliana leaves, the circadian clocks of individual cells may be synchronized through self-organized traveling waves [Wenden et al., 2012; Gould et al., 2018] (Figure 5.7C). While these waves are known to occur through interactions between cells on a regular lattice, the exact interaction mechanism remains unknown [Greenwood et al., 2019]. Finally, in planaria — flatworms that regenerate their bodies after they are cut into pieces — a self-organized

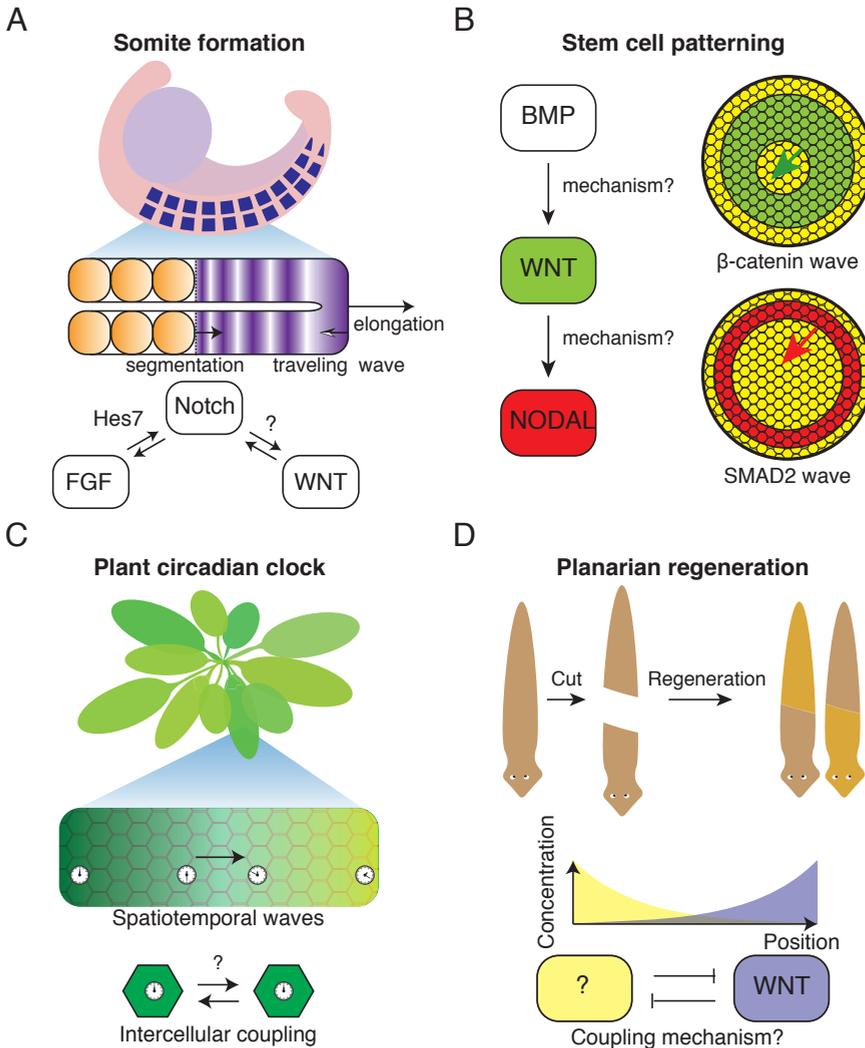
Wnt gradient specifies where the tail reforms after it is cut. Researchers believe that an as-yet unidentified signaling molecule may interact with Wnt in a mutually antagonistic way to indicate where the head should reform after it is excised (Figure 5.7D) [Stückemann et al., 2017]. Thus, our results on two-molecule cellular dialogues may provide insights into this system.

Our work revealed that complex, erratic dynamics is integral to the cellular dialogues enabling dynamic spatial patterns. Researchers have experimentally observed irregular, complex heart beats during ventricular fibrillations [Ten Tusscher & Panfilov, 2006; Qu et al., 2014] and turbulent flows of cytoskeletal fluids [Giomi, 2015] and fluids of Min proteins (MinC, MinD, and MinE) from *E. coli* lysates that form patterns on a petri dish [Halatek & Frey, 2018]. Our work expands this repertoire to include pattern formations through cellular dialogues. Such complex spatial-patterning dynamics may be difficult to observe in experiments because genetic or developmental programs might be triggered and “take over” the pattern-forming dynamics before the cells had enough time to exhibit the kind of prolonged, erratic dynamics that we uncovered here. For example, before a pattern finalizes, some of the cells in the tissue or an embryo may turn on a different developmental program such as those that lead to cavitation in parts of the tissue or some of the cells to collectively migrate. Consequently, cells may not have the time to exhibit the prolonged complex dynamics for a sufficiently long enough time for us to experimentally distinguish it from a short-lived, transient dynamics. Moreover, another experimental challenge to observing the prolonged complex dynamics is that one must measure gene-expression levels of every cell in a tissue or an embryo with sufficiently high temporal and spatial resolutions, and do so continuously for a sufficiently long time. With these difficulties in mind, a plate of natural or synthetic cells that use two-molecule cellular dialogues — rather than a full embryo — may allow us to fully observe the complex dynamics using time-lapse microscopy. It may also be interesting to interpret and analyze our work in the context of complex systems theory [Bar-Yam, 2003]. Doing so may link our findings to those of non-living chemical systems that self-organize patterns [Nicolis & Prigogine, 1977].

## S5.4. METHODS

### S5.4.1. DETAILED DESCRIPTION OF OUR MODEL

In this section, we provide a detailed description of a generalized version of our model, which one can apply to an arbitrary number of diffusing molecules that cells secrete and sense. Our aim here is to concisely summarize the model. For motivations behind the assumptions of our model, please see the main text and our earlier studies [Maire &



**Figure 5.7: Self-organized dynamic-pattern-forming systems with poorly understood interactions that our theoretical framework may help in elucidating.** (A–D) Biological systems with two or more interacting pathways that generate spatiotemporal patterns but whose exact mechanisms and cellular dialogues remain poorly understood. (A) During somitogenesis, a wave of gene-expression states propagates along the anterior-posterior axis of an elongating, pre-somite mesoderm. The conventional view is that this wave is mediated by a coupling between individual oscillators — oscillations in expression levels of Wnt, Notch, and Fgf — and/or by large-scale gradients in the gene-expression levels for those molecules. But how Notch regulates Wnt and vice versa remain questionable while Hes7 is known to mediate the Fgf-Notch interaction (Sonnen et al., 2017). Figure partially adapted from (Oates et al., 2012). (Caption continued on next page.)

**Figure 5.7 (previous page): (B)** Waves of  $\beta$ -catenin (green ring) and Smad2 (red ring) expression-levels propagate in a field of stem cells. Although we know that these waves form due to BMP inducing  $\beta$ -catenin (part of the Wnt pathway) and SMAD2 (part of the NODAL pathway), how exactly these two inductions occur remains poorly understood (Chhabra et al., 2019). **(C)** The circadian clocks of each cell within the leaf of *Arabidopsis Thaliana* are thought to be coupled to each other through an as-yet-unknown mechanism, which is suspected to involve a variety of hormones, sugars, mRNAs and other molecules (Greenwood et al., 2019). **(D)** Planaria regenerate themselves after being cut into two or more pieces. This is thought to rely on mutual antagonism between gradients of Wnt expression (purple) and of an as-yet-unidentified molecule (yellow) (Stückemann et al., 2017). Figure partially adapted from (Stückemann et al., 2017).

Youk, 2015a; Olimpio et al., 2018].

We consider  $N$  cells that communicate through  $l$  distinct, diffusing signaling-molecules that the cells secrete and sense. We first consider cells that are placed on a triangular lattice such that each cell has six nearest neighbors, each at a distance of  $a_0$ . We specify the state of the system — a “system state” — by  $\mathbf{X}(t) = \{X_{k(t)}\}_{k=1}^N$ , where  $X_k = (X_k^{(1)}, \dots, X_k^{(l)})$  is the state of cell  $k$ , which we call a “cell state” for cell  $k$ . In our description below, we will distinguish between the *system state*  $\mathbf{X}$  and the *cell state* of cell  $k$ ,  $X_k$ . Suppose that cell  $k$  secretes a signaling molecule  $i$  ( $1 \leq i \leq l$ ) at a rate  $C_k^{(i)}$ , which is bounded below and above as:  $C_{OFF}^{(i)} \leq C_k^{(i)} \leq C_{ON}^{(i)}$ . Note that we allow for the possibility that the lower and upper bounds on the secretion rate can be different for each signaling molecule. The secretion rate is related to the cell state through the relation

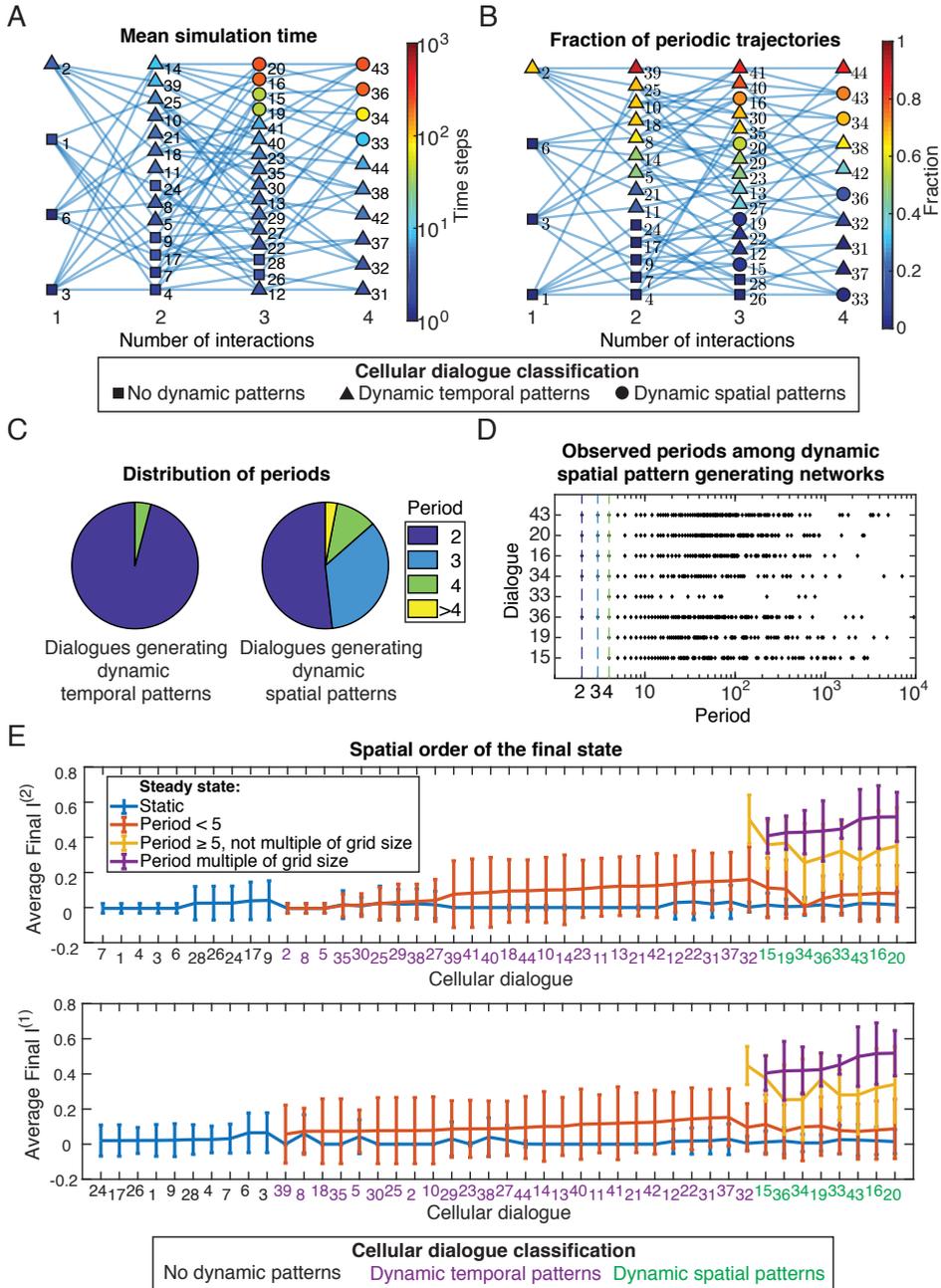
$$C_k^{(i)}(X_k^{(i)}) \equiv (C_{ON}^{(i)} - C_{OFF}^{(i)})X_k^{(i)} + C_{OFF}^{(i)}. \quad (\text{S5.1})$$

In the simplest scenario, the cells secrete signaling molecules at a rate which is either low or high. In this case, each of the  $X_k^{(i)}$  takes binary values — 0 or 1 — such that  $C_k^{(i)}(X_k^{(i)} = 1) = C_{ON}^{(i)}$  and  $C_k^{(i)}(X_k^{(i)} = 0) = C_{OFF}^{(i)}$ . Alternatively, the secretion rate could take continuous values within the closed interval  $[C_{OFF}^{(i)}, C_{ON}^{(i)}]$ . If so, then the cell states are continuous variables (i.e.,  $X_k^{(i)}$  can take any value between 0 and 1). For convenience, we set  $C_{OFF}^{(i)} = 1$  for all  $i$  and measure all concentrations in units of this OFF-secretion rate (which we take to be equal for all molecules, unless we state otherwise).

The concentration of a signaling molecule, once it reaches a steady state, decays with distance from the cell that is secreting it as follows [Olimpio et al., 2018]:

$$\begin{aligned} c^{(i)}(r) &= C_k^{(i)} f^{(i)}(r), \\ f^{(i)}(r) &= \frac{\lambda^{(i)}}{r} \exp\left(\frac{R_{\text{cell}} - r}{\lambda^{(i)}}\right) \sinh\left(\frac{R_{\text{cell}}}{\lambda^{(i)}}\right). \end{aligned} \quad (\text{S5.2})$$

Here we assumed that the cells are spherical with radius  $R_{\text{cell}} \equiv r_{\text{cell}} a_0$  and  $\lambda^{(i)}$  is the



**Figure S5.1: Statistics of trajectories from each of the three classes of cellular dialogues.** (Related to Figure 5.3) All results are based on the same simulation data set also used to generate Figure 5.3E, obtained from Latin hypercube sampling over various system parameters (see Section S5.4.2). (A) Mean simulation time across all simulations of a given network. The simulation time is the time it takes for the system to reach equilibrium or the maximum simulation time if a trajectory never reaches equilibrium. Same graphical representation as in Figure 5.3E. (B) Fraction of trajectories with a periodic final state, i.e. a steady state where the final pattern repeats itself after a fixed number of time steps greater than one. Same graphical representation as in Figure 5.3E. (Caption continued on next page.)

**Figure S5.1 (previous page):** (C) Distribution of the periods of the periodic final states among networks that generate dynamic temporal patterns (Figure 5.3C) and networks that also generate dynamic spatial patterns (Figure 5.3D). (D) Trajectory periods found in simulations of the networks generating dynamic spatial patterns (Figure 5.3D). Each diamond represents a period that was observed in at least one simulation. (E) Average final spatial index for each of the two genes, sorted by cellular dialogue. The trajectories are divided into different classes depending on the final period of the trajectory (represented by differently colored lines). The average is taken over all trajectories within each class. For example, the purple data points show the average values for the subset of trajectories that have a period which is a multiple of the grid size. Error bars represent s.e.m.

5

diffusion length of signaling molecule  $i$ . The diffusion length measures how far the molecule can typically travel before degrading and is set by the molecule's diffusion constant and degradation rate [Olimpio et al., 2018]. Here, we also introduced an “interaction function” function  $f^{(i)}(r)$  to capture the distance-dependent decay. Note that  $C_{ON}^{(i)}$   $C_{OFF}^{(i)}$  are effective secretion rates for ON- and OFF-cells respectively that lump together several terms which appear in the reaction-diffusion equation for molecule  $i$ . They can depend on the diffusion lengths  $\lambda^{(i)}$ . But we will consider them to be independent of the diffusion lengths by assuming that  $C_{ON}^{(i)}$  and  $C_{OFF}^{(i)}$  remain constant as we change  $\lambda^{(i)}$  by tuning other parameters which we do not specify here for brevity. We can reduce the number of parameters by expressing all lengths in units of lattice spacing,  $a_0$ . We define  $l^{(i)} \equiv \frac{\lambda^{(i)}}{a_0}$  and rewrite the interaction function as

$$f^{(i)}(\rho) = \frac{1}{\rho} \exp\left(\frac{r_{\text{cell}} - \rho}{l^{(i)}}\right) \sinh\left(\frac{r_{\text{cell}}}{l^{(i)}}\right), \quad (\text{S5.3})$$

where we introduced  $\rho \equiv \frac{r}{a_0}$ . Here,  $r_{\text{cell}} \equiv R_{\text{cell}}/a_0$  is the radius of a cell expressed in units of the lattice spacing  $a_0$ .

At any given time, the concentration that a cell senses is the sum of the concentrations due to each of the cells in the system. We express the concentration of molecule  $i$  that a cell  $k$  senses as

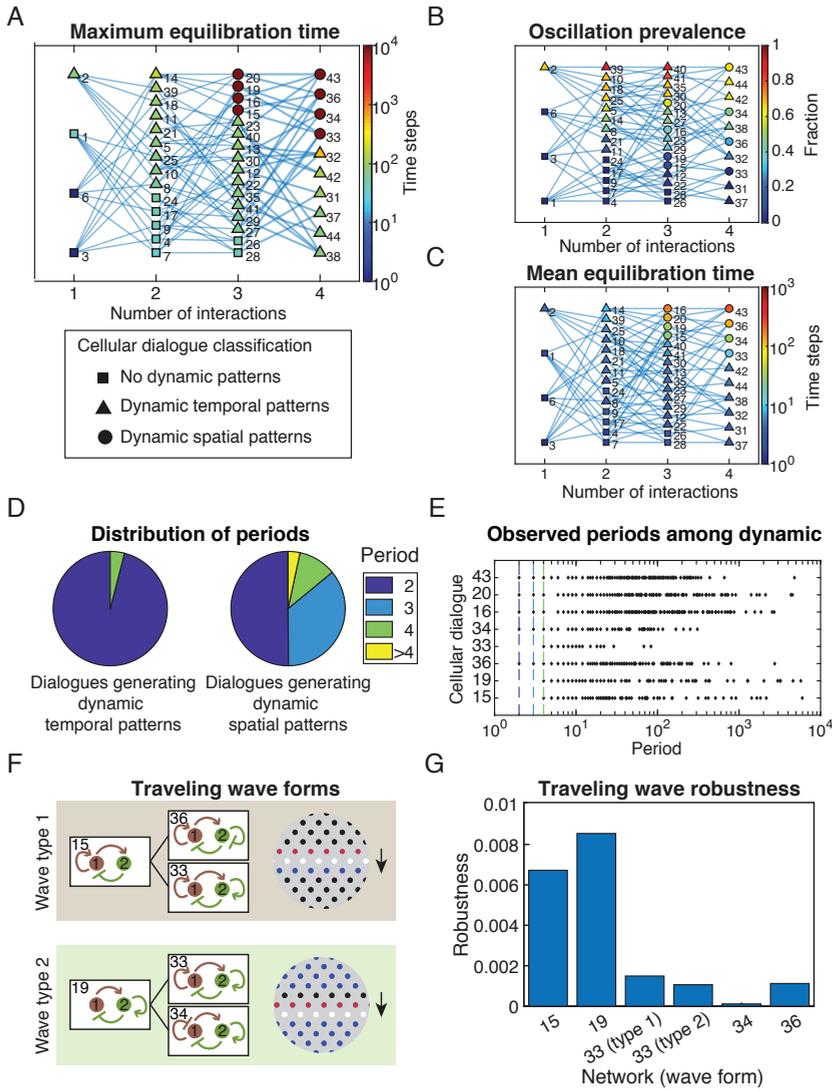
$$Y_k^{(i)} = \sum_{m=1}^N f_{km}^{(i)} C_m^{(i)}, \quad (\text{S5.4})$$

where  $f_{km}^{(i)}$  is a distance-dependent interaction strength between cells  $k$  and  $m$ . Explicitly, we have

$$f_{km}^{(i)} \equiv \begin{cases} f^{(i)}(r_{km}) & (k \neq m) \\ 1 & (k = m) \end{cases}, \quad (\text{S5.5})$$

with  $r_{km}$  being the distance between cells  $k$  and  $m$  and  $f^{(i)}(r)$  as defined in S5.2.

For later reference, we introduce an “interaction strength”  $f_N^{(i)}$  for each signaling molecule



**Figure S5.2: Self-organization of static and dynamic patterns for cellular dialogues with OR-logic signal integration, instead of AND-logic.** (Related to Figure 5.6) Results here can be compared with Figures 5.3B-E. Parameter sets are independently generated from Latin hypercube sampling, with  $10^4$  different parameter sets per cellular dialogue (see Section S5.4.2). (A) By measuring the “equilibration time” - time taken for a simulation to terminate due to either forming a static configuration or a dynamic pattern - for many simulations that all used the same cellular dialogue, we obtained the largest possible (maximum) equilibration time. To ensure that we do not let a simulation run forever, we arbitrarily forced simulations to terminate after  $t_{max} = 10^4$  timesteps if it has not already terminated by itself. (B) Fraction of simulations that exhibit a periodic pattern (i.e., a dynamic pattern that repeats itself after a fixed number of timesteps which is larger than one). (C) Averaging the equilibration times of all simulations for each cellular dialogue. (D) Distribution of periods for periodic patterns. Results shown for cellular dialogues that generate dynamic temporal patterns (Figure 5.3C) and cellular dialogues that also generate dynamic spatial patterns (Figure 5.3D). (Caption continued on next page.)

**Figure S5.2 (previous page):** (E) Distribution of periods for periodic patterns for cellular dialogues that generate dynamic spatial patterns (Figure 5.3D). Each diamond represents a period that we observed in at least one simulation. (F) Wave configurations observed with the OR-logic (compare with the AND-logic shown in Figure 5.4D). (G) Robustness of wave propagation for all six instances of waves shown in (F), defined as the Q-value obtained from simulations (see Figure S5.10 for comparison with the AND-logic).

$i$ :

$$f_N^{(i)} = \sum_{m \neq k} f_{km}^{(i)}. \quad (\text{S5.6})$$

Note that if all cells secrete at the same rate  $C^{(i)}$ , then they would all sense the following concentration

$$Y^{(i)} = (1 + f_N^{(i)})C^{(i)}. \quad (\text{S5.7})$$

#### REGULATORY INTERACTIONS

We now consider how the sensing of one signaling molecule affects the secretion of itself and other molecules by a cell. Molecule  $j$  can affect the secretion of signaling molecule  $i$  in three distinct ways (note that  $i$  can be equal to  $j$ ). First, molecule  $j$  may activate secretion of molecule  $i$ , meaning that a higher concentration of  $j$  leads to a higher secretion rate of  $i$ . Secondly, molecule  $j$  may repress secretion of molecule  $i$ , meaning that a higher concentration of  $j$  leads to a lower secretion rate of  $i$ . Finally, molecule  $j$  may not influence the secretion rate of molecule  $i$  at all. We can capture these three possibilities by an “interaction matrix”  $M_{\text{int}}$ , defined as

$$M_{\text{int}}^{(ij)} = \begin{cases} 1 & j \text{ activates } i \\ -1 & j \text{ represses } i \\ 0 & \text{no interaction between } j \text{ and } i \end{cases} \quad (\text{S5.8})$$

The interaction matrix allows us to define a “cellular dialogue” as a directed (multi)graph in which each node represents one of the  $l$  signaling molecules and each directed edge represents one molecule (node) controlling the secretion rate of either itself (self-loop) or another molecule (directed edge from one node to another) as dictated by the interaction matrix.

A cell may respond in one of multiple possible ways to the sensed concentrations of all the signaling molecules. Its biochemical circuitry sets its response. Here we consider a relatively simple case in which the cell senses the extracellular signaling molecules and then uses one of two standard logic gates — AND and OR gates — to integrate the signals triggered by the sensed molecules to regulate the genes that encode each of the signaling molecules. These gates apply to cells with infinite and finite Hill coefficients (i.e.,

cell's response is not necessarily binary for either logic gates). First, let us consider a cell that uses an AND-gate to integrate the intracellular signals triggered by the sensing of two signaling molecules. For two signaling molecules that activate the secretion of each other, one way to achieve an AND-gate is having two transcription factors — one for each signaling molecule — both needing to bind to the promoter of the gene that encodes the signaling molecules. One of the two transcription factors alone binding to the promoter would be insufficient for activating expression of — and thus secretion of — any of the signaling molecules. Only when both transcription factors are bound to the same promoter, their cooperative interactions would induce the expression and secretion of the signaling molecule that the gene encodes [Buchler et al., 2003]. This scenario leads to a multiplicative update rule for our model. Namely, we determine the cell's secretion rate at the next time step in the cellular automaton by multiplying several mathematical functions — one for each transcription factor — with each function describing the bound fraction of a given transcription factor. Alternatively, a cell may use an OR-gate to regulate genes that encode the two signaling molecules. Here, either of two transcription factors can induce transcription, without the need for both transcription factors to be present. In practice, this can be realized by placing strong binding sites for both molecules at a considerable distance apart, so that the two transcription factors can individually bind to the promoter and recruit RNA polymerases [Buchler et al., 2003].

Mathematically, let  $g^{(ij)}(\mathbf{X})$  be the result of the regulation of the gene that encodes molecule  $i$  by molecule  $j$ , given a system state  $\mathbf{X}$ . If  $g^{(ij)}(\mathbf{X}) = 1$ , then the gene is activated or un-repressed, whereas  $g^{(ij)}(\mathbf{X}) = 0$  means that the gene is either un-activated or repressed. The specific mathematical form of  $g^{(ij)}(\mathbf{X})$  depends on the regulatory interaction. As a general form, we can write it as

$$g_k^{(ij)}(\mathbf{X}(t)) = \theta \left( \left( Y_k^{(j)} - K^{(ij)} \right) M_{\text{int}}^{(ij)} \right), \quad (\text{S5.9})$$

where  $\theta(x) = \begin{cases} 1 & x > 0 \\ 0 & x < 0 \end{cases}$  is the step function. The value of  $\theta(0)$  is unspecified, but to be

consistent with the case of not having a regulatory interaction ( $M_{\text{int}}^{(ij)} = 0$ ), we set  $\theta(0) = 1$  for the AND-logic and  $\theta(0) = 0$  for the OR-logic. Using the standard syntax of Boolean algebra, we can denote the AND-operation as  $\wedge$  and the OR-operation as  $\vee$ . Then, using arithmetic representation of logic gates, we have  $x \wedge y = xy$  and  $x \vee y = x + y - xy$ . Hence, a cell's response with an AND-gate takes the form

$$X_k^{(i)}(t+1) = g_k^{(i1)}(\mathbf{X}(t)) \wedge g_k^{(i2)}(\mathbf{X}(t)) = g_k^{(i1)}(\mathbf{X}(t)) g_k^{(i2)}(\mathbf{X}(t)), \quad (\text{S5.10})$$

and a cell's response with an OR-gate takes the form

$$X_k^{(i)}(t+1) = g_k^{(i1)}(\mathbf{X}(t)) \vee g_k^{(i2)}(\mathbf{X}(t)) = g_k^{(i1)}(\mathbf{X}(t)) + g_k^{(i2)}(\mathbf{X}(t)) - g_k^{(i1)}(\mathbf{X}(t))g_k^{(i2)}(\mathbf{X}(t)), \quad (\text{S5.11})$$

We can readily generalize these expressions to cells with more than two signaling molecules by using the standard rules of Boolean algebra.

#### STEADY STATES OF THE SYSTEM

For regulatory interactions with infinite Hill coefficients, each cell has one of two states for each signaling molecule — OFF (i.e., basally secreting the molecule) and ON (i.e., maximally secreting the molecule). Hence, if the system has a total of  $N$  cells, the total number of possible gene-expression states for the population is finite ( $2^N$ ), meaning that the system (i.e., population) is bound to eventually reach one of two types of steady states in terms of the population-level gene-expression:

1. Stationary steady-state: There is a time  $t^*$  such that for all  $t \geq t^*$ , the system does not change any more (i.e.,  $\mathbf{X}(t+1) = \mathbf{X}(t)$ ). Simply put, this means that the population-level gene-expression state remains constant starting at time  $t^*$ .
2. Periodic steady-state: There exists a time  $t^*$  after which we have  $\mathbf{X}(t+\tau) = \mathbf{X}(t)$  for all  $t \geq t^*$ . Then  $\tau$  is the period of the periodic steady-state. Simply put, this means that the population-level gene-expression state undergoes a periodic oscillation with a period  $\tau$ .

The  $t^*$  — which we will call *equilibration time* — is the time that the system takes to reach either one of the two types of steady states. For stationary steady-states, this is simply the first time when the system reaches a state that does not change over time any more. For periodic steady-states, we define the equilibration time when the onset of the periodicity occurs.

#### ENUMERATING CELLULAR DIALOGUES

If we have two signaling molecules, there are four possible interactions between those two molecules. Each interaction can be either activating, repressing or absent. Hence two molecules can form a total of  $3^4 = 81$  possible cellular dialogues. However, many of these cellular dialogues are equivalent to one another because swapping the labels “1” and “2” on the two molecules (Figure 5.1C) conserves the topology of the graphs that represent the cellular dialogues (i.e., which molecule is labeled “1” or “2” is arbitrary). Under this label-swapping operation, the interaction matrix becomes

$$\begin{pmatrix} M_{\text{int}}^{(11)} & M_{\text{int}}^{(12)} \\ M_{\text{int}}^{(21)} & M_{\text{int}}^{(22)} \end{pmatrix} \mapsto \begin{pmatrix} M_{\text{int}}^{(22)} & M_{\text{int}}^{(21)} \\ M_{\text{int}}^{(12)} & M_{\text{int}}^{(11)} \end{pmatrix}$$

Hence, for cellular dialogues that are invariant under the label-swapping operation, we must have  $M_{\text{int}}^{(11)} = M_{\text{int}}^{(22)}$  and  $M_{\text{int}}^{(12)} = M_{\text{int}}^{(21)}$ , leaving us with two independent elements in the interaction matrix. Each of these two elements can have one of three possible values. Thus, the cellular dialogues that are invariant under the label-swapping operation reduce down to a set of 9 distinct cellular dialogues. We can reduce the remaining 72 cellular dialogues to a set of 36 unique cellular dialogues. Hence, we have total of 45 distinct cellular dialogues. After neglecting the trivial cellular dialogues — those in which neither of the two molecules regulates the other — we obtain the set of 44 cellular dialogues that are shown in Figure 5.3. Note that we also enumerate all cellular dialogues in which a molecule regulates itself but does not regulate the other molecule.

#### POPULATION-LEVEL DESCRIPTION

To characterize the population-level behavior without focusing on the state of every single cell, we introduce “macroscopic variables”. Specifically, we define two macroscopic variables for each molecule, leading to a total of four macroscopic variables for a population. One of them is the average expression level of the gene that encodes molecule  $i$  which, in the case of the digital cells, is equal to the fraction of cells that have gene  $i$  turned on:

$$p^{(i)} = \frac{1}{N} \sum_{k=1}^N X_k^{(i)}. \quad (\text{S5.12})$$

The other macroscopic variable is the “spatial index” for gene  $i$ , which characterizes how spatially correlated the expression levels for gene  $i$  is among the cells. We first introduced this in earlier studies [Maire & Youk, 2015a; Olimpio et al., 2018] and we now define it for a population with multiple signaling molecules as follows:

$$I^{(i)} = \frac{1}{\sum_{n \neq m} f_{mn}^{(i)}} \frac{\sum_m \sum_{n \neq m} f_{mn}^{(i)} (X_m^{(i)} - \langle X^{(i)} \rangle) (X_n^{(i)} - \langle X^{(i)} \rangle)}{\frac{1}{N} \sum_{m=1}^N (X_m^{(i)} - \langle X^{(i)} \rangle)^2} \quad (\text{S5.13})$$

The spatial index for gene  $i$ ,  $I^{(i)}$ , quantifies how spatially ordered the cells are in terms of their expression level for gene  $i$ . It can have a value between -1 and 1, with negative values indicating that neighboring cells tend to have different gene-expression levels (such as in checkerboard patterns or anti-ferromagnetism in spin models) and positive values indicating that neighboring cells that tend to have similar gene-expression levels (forming islands with the same gene expression level, similar to ferromagnetism). When  $I^{(i)} = 0$ , the cells’ expression levels of gene  $i$  are, on average, uncorrelated. In the case of spatially ordered patterns such as traveling waves, the values of  $I^{(i)}$  are positive and relatively high, with exact values depending on the parameters of the system and the wave’s shape.

Together, the set of macroscopic variables  $\{p^{(i)}, I^{(i)}\}_{i=1}^l$  — for  $l$  signaling molecules (we considered  $l = 2$  here) — convey population-level information. However, this description does not contain information about correlations between different genes. For example, we may specify, for two-molecule cellular dialogues, that  $p^{(1)} = p^{(2)} = 0.5$  and  $I^{(1)} = I^{(2)} = 0.5$ . This tells us that half of the genes of each type are turned on and that the cells which have a certain gene on will tend to cluster with other cells that have the same gene turned on. However, we cannot infer whether a cell that has gene 1 turned on is likely to have gene 2 turned on as well or whether its neighbors tend to have gene 2 turned on. There are different ways to consider metrics that also consider such cross-correlations. For example, we can group together cells with each of the four cell states (i.e., (gene 1=ON, gene 2=ON), (ON, OFF), (OFF, ON), (OFF, OFF)) and study the evolution of these populations. However, the disadvantage of this approach is that it does not easily generalize to continuous gene-expression states that we also consider in our work (i.e., for gene regulations with finite Hill coefficients). Alternatively, we can use established statistical metrics for correlations between two sets of values (i.e., gene-expression levels for the two different genes) such as the *Hamming distance*, the *Jaccard index (JI)* and the *Sørensen-Dice* coefficient. As we are mainly interested in knowing whether a spatial configuration is ordered or disordered (i.e., whether the cells have an “interesting” pattern or not), we have not studied such cross-correlations. Nevertheless, our open-source software, MultiCellSim, computes the cross-correlation along with  $p^{(i)}$  and  $I^{(i)}$ .

#### MOVING AVERAGES

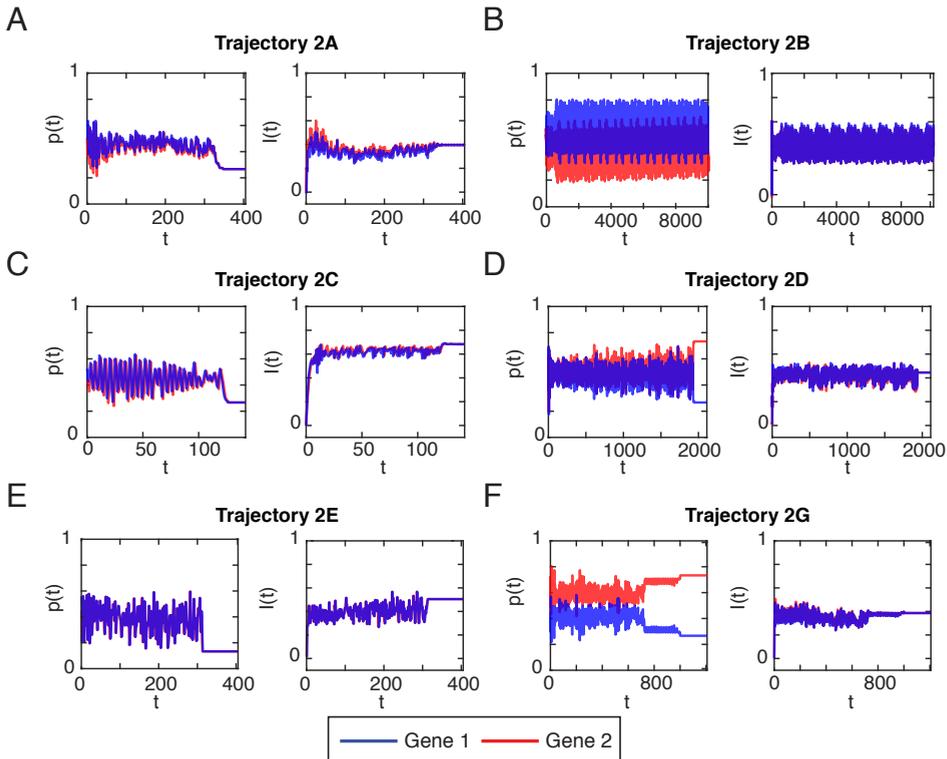
We calculated the Fano factor — variance divided by the mean — for each macroscopic variable in Figures 5.5C and S5.3 by using a sliding time-window of 10 timesteps (i.e., for a macroscopic variable  $y(t)$ , we compute its mean and variance for values of  $t$  within the interval  $(t, t + 10)$ ). Specifically, we calculated a moving variance using the MATLAB function *ttmovvar* and the moving mean using *ttmovmean*. The Fano factor represents a signal-to-noise ratio within a given time-window frame.

### S5.4.2. SIMULATION AND ANALYSIS OF THE MODEL

In this section, we provide a concise overview of our simulations and analyses.

#### FIXING INITIAL CONDITIONS

We started simulations by generating a randomly chosen, initial spatial-configuration that is subject to certain constraints. Unless we chose  $p^{(i)}$  and  $I^{(i)}$  to each have a specific value at the beginning of a simulation, we let each cell to have a 50% chance of having gene  $i$  be ON. This tends to generate spatial configurations in which half of the cells have gene  $i$  turned on. In some cases, we chose  $p^{(i)}$  and  $I^{(i)}$  to each have a specific value at



**Figure S5.3: Additional examples showing how the two macroscopic variables — spatial index and fraction of cells that are ON for each of the two genes — vary over time during self-organization of dynamic spatial patterns. (Related to Figure 5.5)** (A-F) Plots show how the fraction of cells that have a certain gene ON and the “spatial index” for that gene changes over time (see Section S5.4.1). Each panel here corresponds to one of the filmstrips shown in Figure 5.2, as indicated by the panel titles. Time is in units of discrete time steps. Blue curves correspond to gene 2. For each pattern-forming dynamics, we show two graphs corresponding to the graphs of the macroscopic dynamic, such as shown in Figure 5.5B. Left: mean fraction of cells  $p(t)$  that have the indicated gene ON. Right: Spatial index  $I(t)$  for the indicated gene.

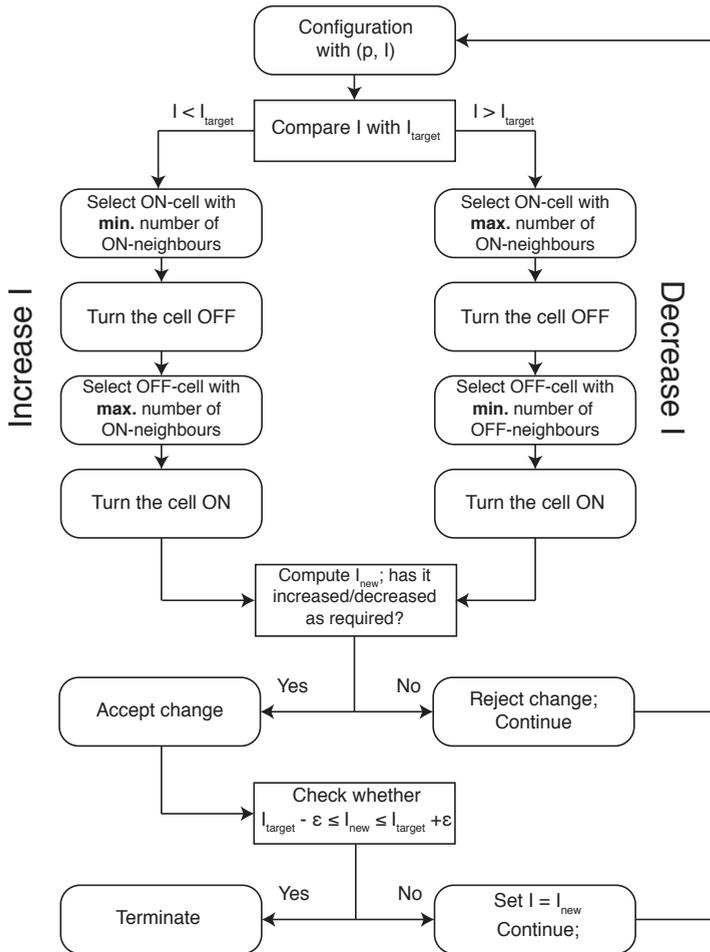
the beginning of a simulation (Figure S5.10). Here, we fixed the value of  $p^{(i)}$  by randomly selecting this fraction of cells, for which we turn on gene  $i$ . To fix the value of  $I^{(i)}$ , we used a Monte Carlo algorithm outlined below.

#### ALGORITHM FOR GENERATING SPATIAL CONFIGURATIONS WITH A GIVEN SPATIAL INDEX

We devised an algorithm that generated spatial configurations, for initializing our simulations, with specified values for the spatial index  $I^{(i)}$  and  $p^{(i)}$ . Our algorithm was motivated by a similar problem in physics — a problem on Ising spin systems — in which one needs to fix the total energy of the spins (analogous to  $I^{(i)}$ ) without changing the average magnetization (analogous to  $p^{(i)}$ ). Our algorithm is illustrated in Figure S5.4 and is as follows:

1. Given a spatial configuration with a given value of  $p$ , start by computing the value of the  $I$  for this configuration.
2. Check whether we should increase or decrease  $I$  by comparing it to the target value  $I_{\text{target}}$ .
3. If  $I < I_{\text{target}}$ 
  - (a) Select the ON-cell with the *minimum* number of neighbors which are also ON. Turn this cell OFF.
  - (b) Select the OFF-cell with the *maximum* number of neighbors which are also OFF. Turn this cell ON.
4. Else if  $I > I_{\text{target}}$ 
  - (a) Select the ON-cell with the *maximum* number of neighbors which are also ON. Turn this cell OFF.
  - (b) Select the OFF-cell with the *minimum* number of neighbors which are also OFF. Turn this cell ON.
5. Compute the spatial index of the new configuration,  $I_{\text{new}}$ . Check whether it has increased or decreased as required.
6. If it has changed as required, accept the change. Go to step 8.
7. Else, reject the new configuration. Go to step 1.
8. If  $I_{\text{new}} \in [I_{\text{target}} - \epsilon, I_{\text{target}} + \epsilon]$ , terminate the simulation.
9. Else, go to step 1 with the new configuration with  $I = I_{\text{new}}$ .

Because we switch the state of both an ON-cell and an OFF-cell, the average number of ON-cells remains constant. To increase  $I$ , we choose cells that tend to have a different state from most of their neighbors and change their state. To decrease  $I$ , we change cells whose state are similar to that of their neighbors. Note that this algorithm is not guaranteed to converge to  $I_{\text{target}}$  because at each iteration of the loop outlined above, we are not guaranteed to increase or decrease  $I$  as required. In particular, if the specified value of  $I$  is outside the range of possible values for  $I$  [Olimpio et al., 2018], the algorithm cannot reach the specified value of  $I$ . Therefore, we typically set a limit on the maximum number of iterations before we terminate the algorithm. Finally, we typically set  $\epsilon = 0.01$ , which allowed for convergence at reasonable speeds while limiting deviations from the target value.



**Figure S5.4: Algorithm used to generate spatial configurations with a desired value of spatial index  $I^{(i)}$  and a fraction of cells which are ON for gene- $i$  (denoted  $p^{(i)}$ ).** (Related to Figure 5.1) Section S5.4.1 defines the two “macrostate” variables,  $I^{(i)}$  (spatial index for gene- $i$ ) and  $p^{(i)}$  (fraction of cells that express gene- $i$  (i.e., ON for gene- $i$ )). Since each cell has two genes — one for each secreted molecule — there are two spatial indices and two  $p$ ’s. Atypical simulation, unless stated otherwise, started with a maximally disordered field of cells (i.e., cells’ gene-expression levels were spatially uncorrelated, for both genes). This means that, at the beginning of a simulation, the spatial index  $I^{(i)}$  was zero for both genes. While keeping the spatial indices to be zero, we could set  $p^{(1)}$  and  $p^{(2)}$  to be virtually any value that we desired, by using the algorithm shown here. In short, the algorithm starts with a spatial configuration described by some pair of macroscopic variables  $(p^{(i)}, I^{(i)})$ , which may not have the values that we want. The algorithm then iteratively updates the value of  $I^{(i)}$  while keeping  $p^{(i)}$  constant. It does so by randomly selecting cells and changing their states until we obtain the  $(p^{(i)}, I^{(i)})$  that we want to begin our simulation with. As  $(p^{(i)}, I^{(i)})$  is specific for a single gene (i.e., gene- $i$ ), we can vary  $(p^{(1)}, I^{(1)})$  independently of  $(p^{(2)}, I^{(2)})$ .

## TERMINATING SIMULATIONS

As noted earlier, we terminate a simulation either when the population reaches a steady state or the simulation reaches the maximum number of time steps  $t_{\max}$ , which we arbitrarily set to be a sufficiently large number. As an example, we chose  $t_{\max} = 10^5$  for populations with  $N = 225$  cells. At each time step, we checked for stationary steady-states by comparing the current system state with the previous timestep's system state. To check for periodic steady-states, one might resort to manually checking the system state at every timestep to see whether the current state has been visited earlier. This becomes computationally infeasible for running many simulations. So, we devised a more efficient scheme for detecting whether the simulation has entered a periodic steady-state. Instead of checking at every timestep, at every  $t_{\text{check}}$  timesteps we manually check whether the previous system state has been visited earlier (we chose  $t_{\text{check}} = 10^3$ ). If we find periodicity in the system states, we ran a second algorithm to find the earliest time at which any state has repeated itself, which told us when the periodic steady-state began.

5

## BATCH SIMULATIONS

Many of the results presented here are from batch simulations, which means that we performed a large set of simulations and obtained statistics on various measures. In many cases, we fixed all parameter values and only varied the initial spatial configuration. By performing a large set of such simulations, we could distinguish whether an observed feature was a general feature for a particular set of parameters or was merely an artifact of a specific initial spatial-configuration. We also ran many simulations, each time varying the parameter values, to find features that were general for a large range of parameter values. Since the parameter values form a continuum, we could not simulate all possible parameter values and thus had to find a way to sample over the space of all parameter values. Specifically, we employed Latin hypercube sampling [McKay et al., 1979], in which we efficiently sampled over a multi-dimensional parameter space by taking parameter sets that were non-overlapping in any of the dimensions. We used this method to generate a large set of conditions for each of the 44 distinct cellular dialogues that we computationally screened. The results in Figures 5.3, S5.1 and S5.2 used this approach. Specifically, to obtain these results, we defined a region in the parameter space in which we varied the parameters  $K^{(ij)}$  and  $C_{ON}^{(j)}$  over a range of values — ranging from 1 to  $10^3$  — while keeping all other parameters held fixed. We sampled parameter values within this region by using the MATLAB function *lhsdesign* to generate a Latin hypercube sample with 10,000 points. We used this approach for each of the 44 cellular dialogues, with both the AND-logic (Figures 5.3 & S5.1) and the OR-logic gate (Figure S5.2).

## IDENTIFYING TRAVELING WAVES

We devised an algorithm for automatically identifying traveling waves in large sets of simulations. Since traveling waves retain their shape while propagating through space, the values of  $p^{(i)}$  and  $I^{(i)}$  would remain constant over time. Due to the periodic boundary conditions that we used, having a traveling wave would mean that the system state returns to itself after  $n$  time steps, where  $n = \sqrt{N}$  and  $N$  is the total number of cells. Hence, we first screened through the simulations to find the ones that had a periodic steady-state with a period that was a multiple of  $n$  — we looked for integer multiples of  $n$  since there may be more complicated waves whose shapes slightly morph as they enter the edges of the field. We next checked whether  $p^{(i)}$  and  $I^{(i)}$  were (sufficiently) constant over the course of one period. Using these two features, we could identify traveling waves in batch simulations without, by eye, examining the simulations explicitly one by one. We then extended the algorithm such that it also gave the orientation of the wave if the cells indeed formed a traveling wave (see Figure S5.5C). We did so through a two-step procedure. First, we distinguished cells that formed the “background” (exterior cells — Figure 5.4A) from the cells that formed the wave band (assuming that there were three states that made up a wave — see Figure 5.4A). Then, we traced the cells from an arbitrarily chosen layer of the wave band to see whether they percolate the system from one horizontal (vertical) edge to the other. If so, then we assigned a horizontal (vertical) orientation to the wave. Else, we assigned a diagonal orientation to the wave.

### S5.4.3. EXTENDING THE MODEL BY ADDING COMPLEX ELEMENTS

In this section, we discuss how we extended our model by adding more complex elements (Figure 5.6A). In the main text and Figures 5.6 and S5.5-S5.13, we describe in detail how adding the four complex elements shown in Figure 5.6A affect the formation and propagation of dynamic spatial-patterns. Below, we dedicate one section for each of these four complex elements as well as for another complex element that is not shown in Figure 5.6 (i.e., spatial gradient of parameter values).

#### STOCHASTIC SENSING AND RESPONSE

There are various sources of stochasticity that can affect pattern formations. These include stochastic expression of the genes that encode the signaling molecules and the fact that a cell cannot determine the concentration of the signaling molecules to an arbitrary level of accuracy (i.e., the Berg-Purcell limit). We did not try to specify the exact source of noise. Instead, we modeled the cells' noisy responses to the signaling molecules by taking a phenomenological approach in which we lumped together various possible sources of stochasticity into a single mathematical term. Specifically, based on our previous work [Olimpio et al., 2018], we let the threshold concentration for how gene  $i$  is regulated by molecule  $j$  — denoted  $K^{(ij)}$  — to fluctuate from cell to cell and from time to time (Figure 5.5A – top left). Mathematically, we can represent this fluctuation as

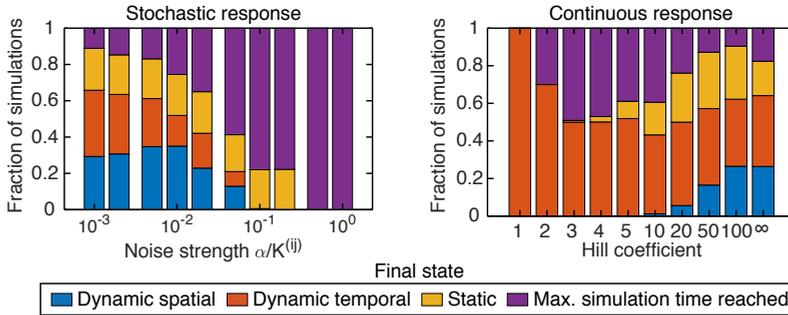
$$K^{(ij)} = K_0^{(ij)} + \delta K^{(ij)}. \quad (\text{S5.14})$$

Here  $K_0^{(ij)}$  is the threshold concentration for gene  $i$  being regulated by molecule  $j$  in the absence of any noise and  $\delta K^{(ij)} \sim N(0, \alpha^{(ij)})$  is a normally distributed random variable. At each time step, we used above equation to update the threshold concentration  $K^{(ij)}$ , independently for each cell. In order to define a global noise-strength without introducing many variables, we have let  $\alpha^{(ij)} = \alpha K_0^{(ij)}$ . In other words,  $\alpha = \alpha^{(ij)} / K_0^{(ij)}$  was fixed for all interactions (i.e., for all pairs  $(i, j)$ ), meaning that the variation in the threshold concentration was proportional to the threshold concentration, with the same proportionality factor  $\alpha$  used for every pair  $(i, j)$ .

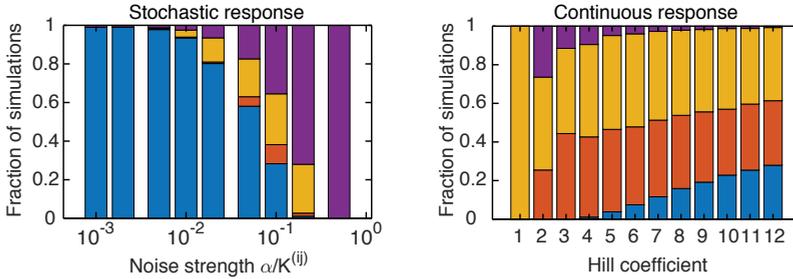
#### CONTINUOUS CELL-RESPONSE FUNCTION

In our model, we so far assumed that cells are binary and secrete signaling molecule  $i$  at either a low, basal rate  $C_{OFF}^{(i)}$  or at a high, maximal rate  $C_{ON}^{(i)}$ . This is a valid assumption whenever the response function is sufficiently ultrasensitive, as discussed in the main text. However, to treat more gradual response functions, we replaced the step-response function (i.e., infinite Hill coefficient) by a continuous, Hill function with a finite Hill coefficient. The Hill coefficient quantifies the steepness of the Hill function. For simplicity, we assumed that all molecules have the same Hill coefficient  $n$ . The update rule for the

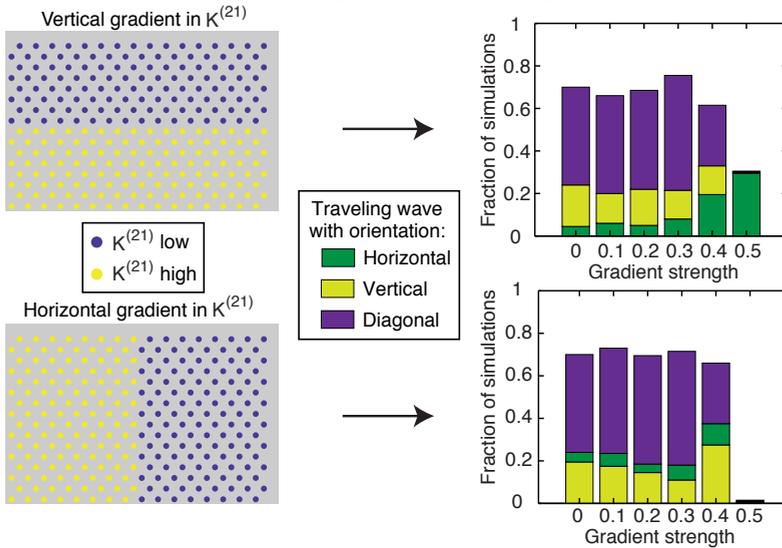
**A Detailed effect of complex elements on self-organization of dynamic patterns**



**B Detailed effect of complex elements on traveling wave propagation**



**C Influence of parameter gradient on traveling wave orientation**



cells' states is still the same as previous, but now with

$$g_k^{(ij)}(\mathbf{X}(t)) = \begin{cases} \frac{(Y_k^{(j)})^n}{(K^{(ij)})^n + (Y_k^{(j)})^n} & \text{activation} \\ \frac{(K^{(ij)})^n}{(K^{(ij)})^n + (Y_k^{(j)})^n} & \text{repression} \\ 1 & \text{no interaction} \end{cases} \quad (\text{S5.15})$$

Note that the Hill coefficient in our model does not have a direct physical interpretation. Instead, it is a phenomenological parameter that describes the steepness of the response function. This is because, in real cells, a ligand-bound receptor typically induces gene expression through a complex signal-transduction cascade rather than through a single molecular process such as a binding of a transcription factor at a promoter. As such, the Hill coefficient does not model any one specific biomolecular process. Therefore, whereas in cooperative binding models, Hill coefficients less than one and larger than two are rare, our model allows for the Hill coefficient to be arbitrarily high or low.

5

#### DISORDERED CELL POSITIONS

In the previous sections, we considered cells to be on a triangular lattice. This is a fair representation of certain multicellular systems (see Table S1 in Olimpico et al., 2018 for a list of examples). But in general, communicating cells do not need to be on a regular lattice. To extend our model to account for alternative spatial arrangements of cells, we adapted our model to allow for randomization of the cell positions through an algorithm adapted from Markov Chain Monte Carlo (MCMC) simulations of hard spheres [Krauth, 2006]. The algorithm allowed us to tune the degree of randomness of the cell positions,

---

**Figure S5.5 (preceding page): Formation and propagation of traveling waves under the influence of noisy gene-expression, finite Hill coefficient, and spatially changing parameter values. (Related to Figure 5.6).** (A-B) Detailed breakdown of simulations with two of the complex elements — stochastic response and continuous response — into four classes of patterns. Dynamic spatial pattern here refers to traveling waves specifically, dynamic temporal patterns to all other simulations that yielded periodic steady states, static patterns to simulations where the final state was non-periodic and max. simulation time reached to simulations that never settled down to a steady state within the total simulation time (10,000 timesteps). (A) Effect of complex elements on the formation of dynamic patterns, corresponding to the data also used in Figure 5.6E (upper panels). We performed 200 simulations for each value of the noise and 150 simulations for each value of Hill coefficient. (B) Effect of complex elements on traveling wave propagation, corresponding to the data also used in Figure 5.6D (upper panels). We performed 1,000 simulations for each value of the noise and 2,534 simulations for each value of Hill coefficient. Each simulation corresponds to a different parameter set for which in the absence of noise and with infinite Hill coefficient a straight traveling wave, such as depicted in the lower panel of Figure 5.6B, can propagate. (C) Effect of a parameter gradient on the orientation of formed traveling waves. Specifically, we considered a step-function gradient for the threshold parameter  $K^{(21)}$  oriented along either the vertical direction (upper panels) or horizontal direction (lower panels). We classified the orientation of the formed traveling waves as the relative gradient strength (see Section S5.4.3) is increased. The unclassified simulations in the bar graphs did not form traveling waves. We performed 200 simulations for each value of the gradient strength.

varying from a perfect lattice to a fully disordered arrangement of cells. However, we still assumed that the cells are immobile or move at a much slower time scale than the time scale involved in molecular/gene-expression changes.

We modeled the cells as 2D hard spheres with a radius of  $R_{\text{cell}}$  (identical for all cells). The cells were placed in such a way that no two cells overlapped. Initially, the cells were placed on a regular hexagonal lattice, with distance  $a_0$  between the cells. We selected a random cell  $j$  with position  $x_j = (x_j^{(1)}, x_j^{(2)})$ . We then performed a Monte Carlo step, where we attempted to move the cell by a displacement,  $x_j \rightarrow x_j + \delta x_j$ . Here  $\delta x_j = (\delta x_j^{(1)}, \delta x_j^{(2)})_j$ , with  $\delta x_j^{(1)}, \delta x_j^{(2)}$  being two random variables that are independent of each other and drawn from a *uniform* distribution on  $[-\varepsilon, \varepsilon]$ . If the cell did not overlap with any other cell at the new position, we accepted the move. Otherwise, we rejected the move and a new move was proposed. To avoid repeated rejections, the cell radius and  $\varepsilon$  were chosen to be sufficiently small. In all our simulations, we took  $R_{\text{cell}} = 0.2a_0$  and  $\varepsilon = (a_0 - 2R_{\text{cell}})/4 = 0.15a_0$ .

The number of Monte Carlo steps we performed using this algorithm is a measure for the degree of randomness in our cells' positions. As a rough indication, for a system of  $N = 144$  cells, after 100 Monte Carlo steps, the arrangement still appears to be very similar to a minutely perturbed lattice. After  $10^4$  Monte Carlo steps, we observed that the cells were clearly not on a lattice anymore and that distinct rows and columns of cells were still recognizable. After  $10^5$  Monte Carlo steps, we found that the arrangement of cells looked similar to what one would obtain by randomly "dropping" cells onto a plane. We can make these statements more precise by looking at the spatial distribution of cells surrounding each cell. Quantitatively, we now have a different interaction strength  $f_N^{(i)}$  for each cell in the system. As the cells become more randomly arranged, the distribution of the interaction strengths becomes broader and the mean also increases. From these calculations, one can show for example, that after  $\sim 10^5$  Monte Carlo steps, a field of  $N = 144$  cells obtains spatial configuration that is indistinguishable from that of a field of randomly placed cells.

#### CELL MOTILITY

We also extended our model to account for undirected, diffusive movements of the cells. Researchers have considered diffusive cell motility as a stochastic process and modeled it with the Langevin equation. They have applied this approach to treat fibroblasts [Dunn & Brown, 1987], endothelial cells [Stokes et al., 1991] and granulocytes [Schienbein & Gruler, 1993]. More precisely, these earlier studies have proposed that the underlying process is that of an Ornstein-Uhlenbeck process, in which cells randomly drift while experiencing a restorative force — this force represents friction in a Brownian motion — which tends to bring the cells back to their original positions. A previous work

showed that one can represent the discrete time process that corresponds to this process with the following equation [Dunn & Brown, 1987]:

$$dx(t) = \phi dx(t-1) + \eta(t) + \theta \eta(t-1), \quad (\text{S5.16})$$

where  $dx(t) = x(t) - x(t-1)$  is the displacement of a cell at time  $t$ ,  $\eta(t)$  is a discrete random noise term with mean zero, and  $\phi$  and  $\theta$  are real numbers that depend on the restorative force's strength.

For our system, we took a simpler approach to model cell motility by neglecting the temporal correlations which arise in the frictional term. Hence, we assumed that the cells drift around without one cell's motion being correlated with another cell's motion, as in a classic random-walk and Wiener process. To model cell motility then, we used the same Monte Carlo algorithm that we used for randomizing the cell positions but now move all cells at each time step instead of perturbing the initial positions for a fixed number of cells. We defined the *cell motility*  $\sigma_D$  to be the width of the Gaussian term, in units of  $a_0$ , describing the Brownian motion process through which we update the cell positions. Explicitly, at each time step, we updated each of the  $N$  cells one by one through

$$x_j(t) \rightarrow x_j(t) + \delta x, \quad \delta x \sim \mathcal{N}(0, \sigma_D a_0). \quad (\text{S5.17})$$

Here  $\sigma_D$  is a parameter that quantifies the extent of a cell's motion in units of the lattice constant  $a_0$  (i.e., the distance between two neighboring cells when placed on a regular lattice).

#### SPATIAL GRADIENT OF PARAMETER VALUES

Studies of the Turing-patterning mechanism have revealed that a spatial gradient of production rates and other parameters as well as more complex, spatially anisotropic parameter values can affect in which direction stripes become aligned after forming through Turing instability [Hiscock & Megason, 2015]. Motivated by this observation, we wondered whether spatial gradients of parameter values can influence the direction in which waves would travel after forming in our system. To this end, we experimented with applying spatial gradient of parameter values in various directions and for various parameters. As an example, cells at the top edge may have a higher maximal secretion rate for molecule  $i$  than the cells at the bottom edge, with the maximal secretion rate continuously changing as we traverse the field of cells row by row. Starting from a parameter set which is able to generate waves, we modified one of the parameters  $P$  of a cell  $k$  to be position dependent,

$$P(x_k) = (1 + f(x_k)) P_0, \quad (\text{S5.18})$$

position of cell  $k$ ,  $f(x_k)$  is a modulation term that adjusts the parameter  $PP$ , and  $P_0$  is a constant. The simplest type of a spatial gradient that we could consider was a step function defined in either a horizontal or vertical direction (Figure S5.5C). For example, we could take a vertical gradient by defining  $f(x_j) = A_y \theta(x_j^{(2)})$ , with a step function

$$\theta(x) = \begin{cases} 1 & \text{if } x \geq 0 \\ -1 & \text{if } x < 0 \end{cases} \quad (\text{this assumes that half of the cells are at } x^{(2)} > 0).$$

We then quantified the sharpness of the gradient by a gradient-strength parameter  $A_y$ , which represents the fractional change in the value of  $P$  on either side of the step. Note that with this gradient, the average value of the parameter remains unchanged from cell to cell (i.e.,  $\sum_j f(x_j) = 0$ ).

## S5.5. SUPPLEMENTARY INFORMATION

### S5.5.1. OVERVIEW OF SELF-ORGANIZED PATTERNS

In this subsection, we provide a detailed but mostly qualitative overview of the different types of self-organized patterns we observed in simulations of our model. The aim is give a general overview of the different possible morphologies and dynamic features of these patterns, and to understand basic features of these patterns in terms of the concepts we have introduced or will introduce. In subsection S5.5.3, we will analyze traveling waves — a subset of dynamic spatial patterns — more closely.

#### STATIC PATTERNS

While the focus of this work is mainly on dynamic patterns, we also observed static patterns with a high degree of spatial organization in all two-gene networks studied. They commonly arise after a relatively short transient phase (10s-100s of time steps) and can have different shapes and compositions of cell states (i.e. the cells can have different combinations of gene expression). In terms of shape, most patterns consist of one or more islands or stripes of cells with a different cell state from the surrounding cells. Since there is no Turing mechanism in our system, we did not identify a natural length scale for the patterns, although changing parameters did seem to affect the size of structures such as islands. Patterns were most commonly observed to have two sets of cell states, where one group of cells has one cell state and the other group has another cell state. Patterns with three cell states are rare, but not impossible to generate. We did not observe any patterns where all four cell states existed concurrently. The most common static patterns are ones that also arise in the model with one signaling molecule and consist of one group of cells with a given gene ON and another group with that gene OFF. In the case of two molecules, it is common to find islands with both genes ON (or OFF) with the rest of the system consisting of cells with both genes OFF (or ON). We also observed patterns where the two genes were mutually exclusive, i.e. if a cell has gene 1 ON, it has gene 2 OFF, and vice versa. Finally, we occasionally found a boundary layer separating an region with similar gene expression (e.g. island or stripe) from the rest of the cells with a different cell state than either the region or the rest of the cells.

#### DYNAMIC TEMPORAL PATTERNS

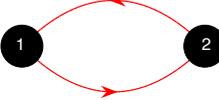
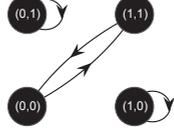
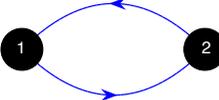
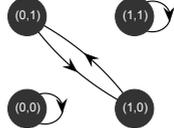
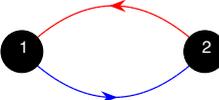
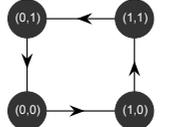
Dynamic temporal patterns are periodic steady states where the system returns to an earlier state after a finite number of time steps  $\tau > 1$  (the period of the oscillation), but do not propagate information across space.

**Single-cell oscillations** Oscillations can arise at the single-cell level in the case of one gene with a negative feedback loop. If certain parameter constraints are satisfied, the

gene expression level of a single cell oscillates between ON and OFF indefinitely. These oscillations are the result of our adiabatic description, where we assume that cells respond slowly compared to the time for signaling molecule concentrations to reach a steady state. An ON-cell turns OFF because the concentration it senses is high enough to suddenly switch to the other state. The OFF-cell then senses a low concentration and the cell switches to the other extreme immediately, without ever reaching the intermediate steady state.

With two genes, oscillations at the single cell level remain relatively simple and can only have periods up to four, since there are only four cell states with two genes. In practice, by examining all possible single-cell state diagrams (Section S5.5.2), we found that the vast majority of single-cell oscillations were of period 2 (see Figure S5.1C). Single-cell period 2 oscillations arise in all networks that can generate dynamic patterns (temporal or spatial), while period 4 oscillations arise only in networks with an incoherent mutual feedback (i.e., for all networks generating dynamic spatial patterns in Figure 5.3D as well as Network 14 in Figure 5.3C).

We can interpret these results by looking at the three core network structure that give rise to most dynamic patterns (Table S5.1). For each of the motifs, the interpretation of the oscillations is straightforward. For mutual repression (Table S5.1 – top row), a cell is able to oscillate between  $(0,0)$  and  $(1,1)$ , whereas  $(0,1)$  and  $(1,0)$  are stable states. When both genes are off, both are unrepressed and will turn on the next time step, after which they are both repressed and turn off again. However, if only one of the genes is on, it represses the other gene but is not repressed itself. For mutual activation (Table S5.1 – middle row), the oscillation is between the states  $(0,1)$  and  $(1,0)$ . Each gene can turn on the other, but turns off when the other gene is on. However, if both genes are ON, they sustained each other, whereas if neither is ON, they also cannot turn ON. Finally, for a positive-negative loop (Table S5.1 – bottom row), the system undergoes a period 4 oscillation between the four states. These results obviously depend on the parameters chosen, but it is intuitive that they should be possible for some set of parameters. Again, these results rely on the separation of time scales between the relaxation of the signaling molecule concentrations and the response of the cells.

Name	Network topology	Single-cell state diagram
Coherent mutual feedback (-/-)		
Coherent mutual feedback (+/+)		
Incoherent mutual feedback (+/-)		

**Table S5.1:** Two gene network motifs generating oscillations. The three core topologies for mutual interaction between the two genes are shown together with typical single-cell state diagrams showing oscillations. The state diagrams are for the case when the concentration of the regulator genes always surpasses the threshold when the gene is ON and is below the threshold when it is OFF, i.e.  $C_{ON}^{(j)} > K^{(ij)} > C_{OFF}^{(j)}$  for all genes  $i, j$ .

5

**Synchronization of single-cell oscillations** Oscillations persist on a multicellular level, with cells synchronizing their oscillations depending on how strongly the cells interact. The degree of synchronization is reflected in the proportion of cells that oscillates together, as well as how they are spatially arranged (e.g. we would consider an oscillating island more synchronized than randomly distributed cells that oscillate in sync). Generally, the oscillations in the multicellular system can vary between completely autonomous (i.e., each cell independently oscillates) to completely synchronized (i.e., all cells oscillate synchronously; see Figure 5.2I). Full autonomy is reached if and only if each of the interactions is in the autonomous (A01) phase (see Section S5.5.2). Full synchronization can be reached for a variety of other parameter conditions. In between, the system can partially synchronize and exhibit domains of cells oscillating together that do not extend over the entire lattice. Oscillatory cells can also co-exist with stationary cells that are in one of the stationary states (for negative-negative or positive-positive feedback).

**Complex dynamic temporal patterns** We distinguish between “simple” oscillations, which are superpositions of single-cell oscillations and thus have a period of at most four, and more complex oscillations, which we will describe here. Oscillations of a more

complicated form arise in the networks that are capable of generating dynamic spatial patterns. Each of these networks produces a wide range of periods with  $\tau \geq 5$  (Figure S5.1D), most of which correspond to dynamic temporal patterns. Typically, they feature oscillating domains that coexist with a background of static cells (e.g., the oscillating island in Figure 5.2H), but where different cells in the domain undergo different cycles of gene expression states over time. This can give rise to complicated temporal patterns because the oscillatory sequences of individual cells may not line up, especially when they are incommensurable.

These complex periods are indeed associated with dynamic patterns, as we can verify by measuring their degree of spatial order using the spatial index  $I^{(i)}$ . Overall, steady states with a period  $\tau \geq 5$  tend to have higher values of  $I^{(i)}$  for both genes (yellow and purple lines in Figure S5.1E), indicating that they tend to be more spatially ordered than oscillations with simple periods (red lines in Figure S5.1E) and static patterns (blue lines in Figure S5.1E).

#### DYNAMIC SPATIAL PATTERNS

Dynamic spatial patterns are characterized by gene expression profiles that translate across space, thereby allowing propagation of information across the multicellular system. These can be rigid profiles of gene expression that move across the system without changing shape, but we also count patterns that move and morph (i.e. change shape) at the same time as dynamic spatial patterns. Note that these patterns require periodic boundary conditions to be sustained indefinitely.

**Traveling waves** Traveling waves are characterized by stripes of cells that translate across the lattice in a regular fashion (see Figure 5.2A, 5.2C-E and 5.2G for examples). They typically consist of three types of cells (with different states) and travel on a background consisting of cells of the fourth type. When two traveling waves in opposite directions collide, they typically annihilate each other, leaving a void of cells with the background state. Characteristics of traveling waves and their propagation conditions will be discussed in subsection S5.5.3.

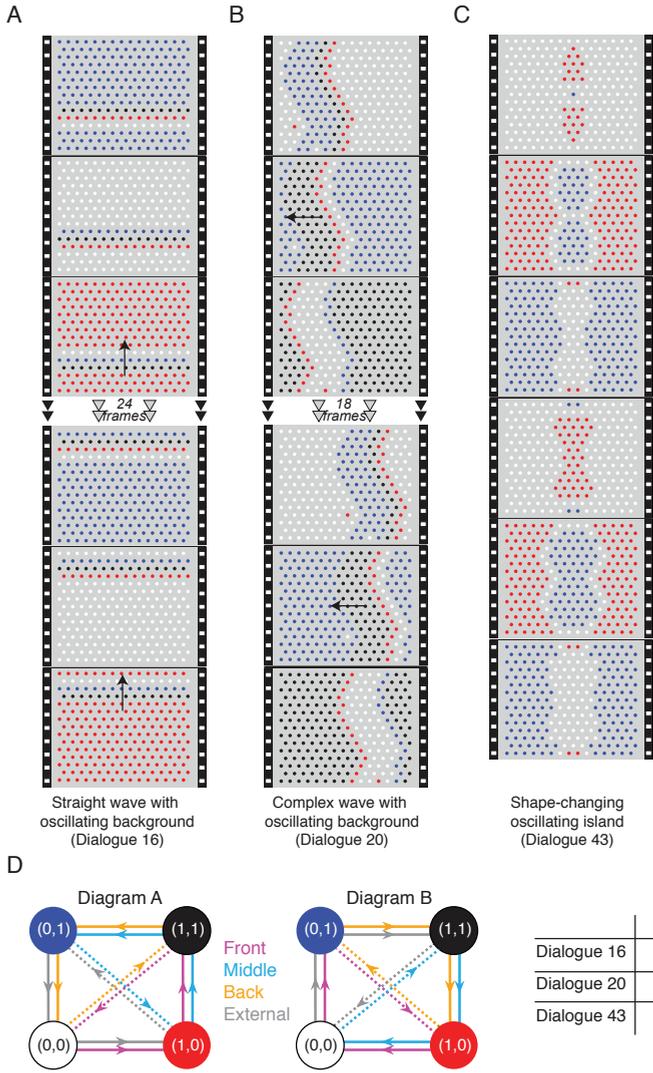
**Complex wavelets** In a number of cases, we observed complex wavelets that propagate indefinitely without repeating themselves, within the maximum simulation time (Figure 5.2B). Since the total number of system states is limited to  $2^N$ , these waves will eventually settle down to a steady state. The transient wave patterns they generate look very similar to (less coordinated) traveling waves, and arise as transient states during the generation of all types of dynamic spatial patterns described here.

**Spiral and concentric waves** Spiral and concentric waves are similar to traveling waves with the main difference that their orientation is outward from a source or center rather than linear in a fixed direction (Figure 5.2F). Locally, they typically look like traveling waves, with the same set of cell states as in traveling waves. Due to annihilation of colliding waves, spiral and concentric waves are less stable, since only particular configurations where the outcome of the collision is an earlier spiral wave pattern will be observed as persistent spiral waves. It is more common to observe spirals and concentric waves as transient patterns that are created and annihilated repeatedly, until the system settles down to a more stable configuration such as a traveling wave.

**Traveling pulses** We also observe small, localized patterns of a few cells that translate across the lattice in a regular way. They are similar to traveling waves, but the traveling pulses are small, localized patterns that do not span the entire size of the system.

## 5

**Oscillatory traveling waves** In networks 16, 20 and 43 — characterized by the incoherent mutual feedback motif without positive self-regulations (Figure 5.3D) — we found oscillatory traveling waves where both the wave states and the background state oscillate over time (Examples in Figure S5.6). At any given fixed time, these waves typically look similar to the non-oscillatory traveling wave, but due to the oscillations the dynamics is different. Perfectly aligned waves where each wave state occupies a single band of cells are relatively rare. Most waves have bands that occupy the width of more than one cell (see for instance Figure S5.6B). The waves undergo a successive sequence of static oscillations followed by an translation (Figure S5.6D). Details of their dynamics are further discussed in Section S5.5.3.



**Figure S5.6: Dynamic patterns with oscillating cells and their associated state diagrams. (Related to Figure 5.3)** (A-C) Examples of dynamic spatial patterns with oscillating cells (in blue) in their background, generated by cellular dialogues 16, 20, and 43 (see Figure 5.3D). Time progresses vertically downwards, with subsequent frames separated by one timestep unless indicated otherwise. (D) State diagrams showing all possible ways that a cell's gene-expression can change over time, for cellular dialogues 16, 20, and 43 shown in (A-C) (see Section S5.5.2). Every cell shown in the filmstrips of (A-C) cycle through three different states before the pattern moves to the next row of cells. The transitions between these single-cell states (nodes in the graphs) are depicted in the state diagrams as directed cycles of a graph. Different colors indicate different relative positions of the cells (matching the colors in Figure 5.4A). One of these transitions — indicated by the dashed lines — is concurrent with the displacement of the pattern. There are two possible state diagrams for the three cellular dialogues that generate oscillatory dynamic spatial-patterns, as indicated in this table.

### S5.5.2. PARAMETER-DERIVED GENERAL CONSTRAINTS ON THE DYNAMICS

In this subsection, we derive a number of methods to derive general constraints on the dynamics of the system from the parameters of that system. First, we introduce the concept of dynamical phase for each regulatory interaction between two (possibly identical) signaling molecules. Next, we introduce the concept of state diagram — a graphical way to represent transitions between cell states — and discuss their usefulness in deducing constraints on the system’s dynamics. We then derive general constraints on the dynamics of a system with multiple signaling molecules, which arise as special combinations of these phases. Finally, we present a formal algorithm to calculate the dynamical constraints and represent them in a state diagram for any set of arbitrary parameters.

#### DYNAMICAL PHASES FOR EACH INTERACTION

The idea behind dynamical phases is based on the observation that for extreme parameter values, the behavior of the system becomes predictable. For instance, if the interaction between cells is very strong and the threshold values characterizing their response are very low, then we expect the cells to always exceed these thresholds regardless of the precise states of the cells. These ideas were made precise in our previous work for systems with one signaling molecule, which represented these dynamical constraints as “phenotype functions” on a “phenotype diagram” [Maire & Youk, 2015a]. In this work, we extend this formalism to a more general framework applicable to multiple interactions.

Consider an interaction between two genes where a regulating gene  $j$  controls the expression of the regulated gene  $i$  (possibly  $i = j$ ). The interaction is specified by the threshold  $K^{(ij)}$  for turning the gene ON/OFF (depending on whether the interaction is activating or repressive), and the ON-secretion rate  $C_{ON}^{(j)}$ . Suppose the cells have an effective distance  $a_0$  to their nearest neighbors, and lead to an interaction strength  $f_N^{(j)}(a_0)$ . Recall that the outcome of the interaction is specified by  $g^{(ij)}(\mathbf{X})$ . We then distinguish the following phases:

1. **P1**: sensed concentration permanently above threshold. The phase is defined by  $(1 + f_N^{(j)})C_{OFF}^{(j)} > K^{(ij)}$ . For an activating interaction, this implies that  $g^{(ij)}(\mathbf{X}) = 1$  for any system state  $\mathbf{X}$ . The interaction is always ON. For a repressive interaction, we have  $g^{(ij)}(\mathbf{X}) = 0$  and the interaction is always OFF. For a single activating interaction, this corresponds to the *all ON phase* — all cells in the system turn ON in one time step and remain ON.
2. **P0**: sensed concentration permanently below threshold. The phase is defined by  $(1 + f_N^{(j)})C_{ON}^{(j)} < K^{(ij)}$ . For an activating interaction this implies  $g^{(ij)}(\mathbf{X}) = 0$  and for a repressive interaction  $g^{(ij)}(\mathbf{X}) = 1$ . For a single activating interaction, this

corresponds to the *all OFF phase* — all cells in the system turn OFF in one time step and remain OFF.

3. **A1**: autonomy whenever the regulating gene is ON. The phase is defined by  $C_{ON}^{(j)} + f_N^{(j)} C_{OFF}^{(j)} > K^{(ij)}$ . For an activating interaction, this implies that  $g_k^{(ij)}(\mathbf{X}) = 1$  whenever  $X_k^{(j)} = 1$ . This means that the interaction is always ON in a cell  $k$  whenever gene  $j$  is ON regardless of the rest of the cells. However, when gene  $j$  is OFF, whether the interaction will be ON depends on the state of other cells. For repression,  $g_k^{(ij)}(\mathbf{X}) = 0$  whenever  $X_k^{(j)} = 1$ . In this case, the interaction is always OFF in cell  $k$  whenever it has gene  $j$  ON. For a single activating interaction, this corresponds to the *activation phase*, whereby ON cells remain ON regardless of the rest of the system.
4. **A0**: autonomy whenever the regulating gene is OFF. The phase is defined by  $C_{OFF}^{(j)} + f_N^{(j)} C_{ON}^{(j)} < K^{(ij)}$ . This is analogous to the previous case with some roles switches. For activation, we get  $g_k^{(ij)}(\mathbf{X}) = 0$  whenever  $X_k^{(j)} = 0$ . For repression,  $g_k^{(ij)}(\mathbf{X}) = 1$  whenever  $X_k^{(j)} = 0$ . For a single activating interaction, this corresponds to the *deactivation phase*, whereby OFF cells remain OFF regardless of the rest of the system.
5. **A01**: autonomy regardless of whether the regulating gene is ON or OFF. This phase is defined by parameter values for which both inequalities of A1 and A0 hold. These conditions can only be met simultaneously if  $f_N^{(j)} < 1$ . For activation, it implies that  $g_k^{(ij)}(\mathbf{X}) = X_k^{(j)}$ . More explicitly, it means that  $g_k^{(ij)}(\mathbf{X}) = 1$  whenever  $X_k^{(j)} = 1$  and  $g_k^{(ij)}(\mathbf{X}) = 0$  whenever  $X_k^{(j)} = 0$ . Hence  $X_k^{(j)}$  fully determines fate of the interaction. For repression, the roles are reversed and  $g_k^{(ij)}(\mathbf{X}) = 1 - X_k^{(j)}$ . For a single activating interaction, this corresponds to the *autonomy phase* — ON cells remain ON and OFF cells remain OFF.
6. **U**: unconstrained. This phase is defined by parameter values for which the conditions of neither A0 nor A1 are true. Hence, we have  $C_{ON}^{(j)} + f_N^{(j)} C_{OFF}^{(j)} < K^{(ij)}$  and  $C_{OFF}^{(j)} + f_N^{(j)} C_{ON}^{(j)} > K^{(ij)}$ . These conditions can only be met simultaneously if  $f_N^{(j)} > 1$ . In this phase, we cannot deduce any general constraints on  $g_k^{(ij)}(\mathbf{X})$  and have to look at the specific system state  $\mathbf{X}$  to determine whether an interaction will be ON or OFF. For a single activating interaction, this corresponds to the *activation-deactivation phase*, whereby cells can both activate (turn ON) as well as deactivate (turn OFF) other cells.

The interpretation of these phases are best understood for a system with only one signaling molecule. For now, note that the phases **P0** and **P1** make the interaction trivial — the outcome is always known regardless of the state of the cell itself or its neighbors. The phases **A0**, **A1** and **A01** place constraints which are dependent on the current state

of the cell, and the phase **U** does not place any constraints on the system's dynamics. For a system with multiple signaling molecules, each interaction will be characterized by one phase. The obvious next question is how to put together the constraints from these different interactions to derive general constraints on the system's dynamics.

For two molecules, we can represent the phase regions and phases of each interaction together on a “phenotype diagram” consisting of two plots, as shown in Figure S5.7C. In this example, each of the interactions lies in a colored region representing the phase. Specifically, the activation  $1 \leftarrow 2$  is in the A0 phase (lower figure, circle with “1”). As this is the only regulatory interaction on gene 1, we now immediately deduce features of the dynamics of gene 1. Specifically, we now know that whenever gene 2 is OFF in a cell, gene 1 will always be inactivated (OFF) the next time step. Similarly, the repression  $1 \dashv 2$  is in the A01 phase, meaning that if gene 1 is ON, then gene 2 will always be repressed. However, if gene 1 is OFF, then it does not repress gene 2, but whether the latter turns on depends also on the result of its own interaction. More generally, we can combine all these phase constraints into general constraints on the dynamics of the system. In the next subsection, we consider limits where the phase constraints completely constrain each cell's dynamics.

## 5

## STATE DIAGRAMS

The basic idea of the state diagram is that it displays all the possible transitions between different cell states of a system. The concept of state diagram has been explored in earlier work on modeling genetic circuits with binary expression states [Thomas, 1990], but has been limited to models of gene circuits at the single-cell level. In our case, the cell states are the binary states  $X = (X^{(1)}, \dots, X^{(l)}) \in \{0, 1\}^l$  specifying for a given cell whether each gene is ON or OFF. For one signaling molecule, there are only two cell states, 0 and 1. For two signaling molecules, we distinguish the four cell states (0, 0), (0, 1), (1, 0) and (1, 1). If a cell in a given cell state  $X$  can adopt the state  $Y$  after one time step — either under its own influence or through sensing molecules secreted by other cells — then we draw an arrow between the states  $X$  and  $Y$ .

For a single interaction, the procedure is straightforward and the diagrams are simple to interpret (see Figure S5.7A). For instance, the P1 phase with activation means that cells always turn ON after one time step, hence giving the diagram with all arrows going to the 1 state and no other arrows. For A01 with an activating interaction, the cells are autonomous, so 0 remains 0 and 1 remains 1. However, A01 with repressive interaction gives oscillations between 0 and 1. An ON-cell always turns OFF because it will always sense a concentration above the threshold, repressing its gene expression, while

an OFF-cell will always turn ON because its sensed concentration will always be below the threshold, leaving the gene unrepressed. The collective set of possible transitions is then displayed as a diagram with the cell states and possible transitions.

For two genes, the interpretation is analogous. We draw a directed graph with the four cell states  $(0, 0)$ ,  $(0, 1)$ ,  $(1, 0)$  and  $(1, 1)$  as nodes and directed edges between these states to indicate (possible) transitions between the states. In Figure 5.4E, in the left diagram for dialogue 14, the transitions are heavily constrained. Each of the four states leads only to one other possible state. This completely constrains the dynamics of the system, so that it becomes completely predictable — any cell's dynamics in the system is completely determined by the transitions in one diagram. As such, this state diagram does not allow for multicellular pattern formation, as all individual cells will oscillate individually. In contrast, if we take a diagram such as the one depicted for dialogue 15 (Figure 5.4E), the dynamics of the system is not entirely constrained. The state  $(0, 0)$  has two arrows leaving from it, indicating that either transition is possible, depending on the exact concentration a cell senses. As such, it allows for a pattern such as a traveling wave to propagate, because cells of the same state do not always evolve in the same way, but evolve depending on the other cells in the system. In practice, the state diagram of dialogue 14 (Figure 5.4E – left figure) can be realized under a diverse range of parameter conditions, whereas the diagrams for dialogues 15 and 19 (Figure 5.4E – right figure) are never realized without additional transitions that we have not drawn in the diagrams here. Note that without additional knowledge, traveling waves are in principle compatible with all diagrams containing the diagrams in Figure 5.4E – right figure as subdiagram. However, the wave can propagate if and only if the particular wave states that we identified (Figure 5.4 and subsection S5.5.3) follow the transitions depicted in the figure. This is not information contained in the diagram itself and must be obtained from further examination and calculation — the state diagrams merely represent the set of possible cell state transitions given knowledge of the parameter of the system.

Two properties of the system are immediately evident from the graphical representation of the state diagram. To begin with, a state diagram tells us which cell states could potentially be stationary. Such states must have an arrow to themselves in the state diagram, which we call a *self-transition*. Should the system reach a non-oscillatory steady state, then that final state can only be composed of cell states which have a self-transition. If there are no self-transitions, then the system cannot generate stationary patterns — this happens for instance in dialogue 14 with certain parameters, which can produce the state diagram shown in Figure 5.4E. If there is only one self-transition, then the only possible stationary steady state is a uniform system where all cells have the state with

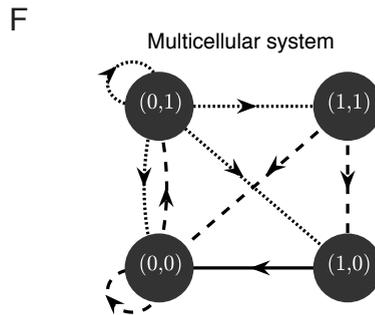
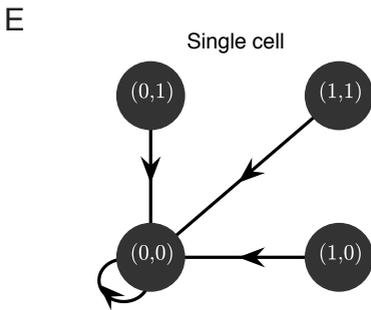
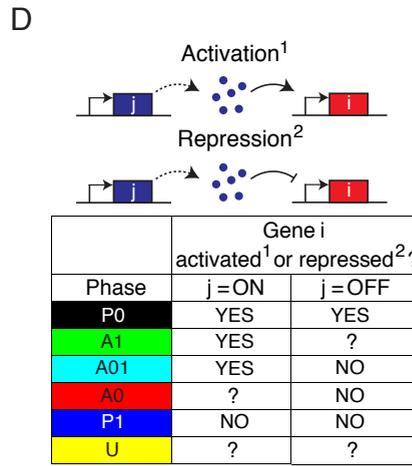
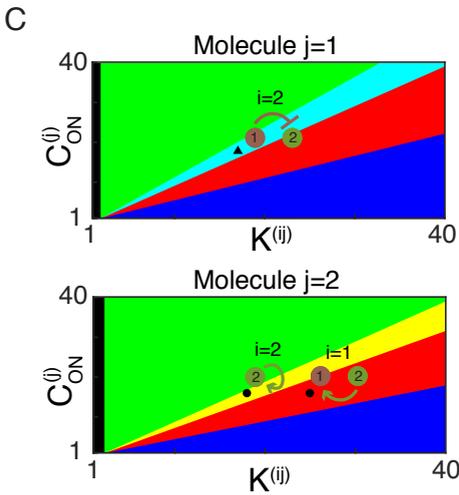
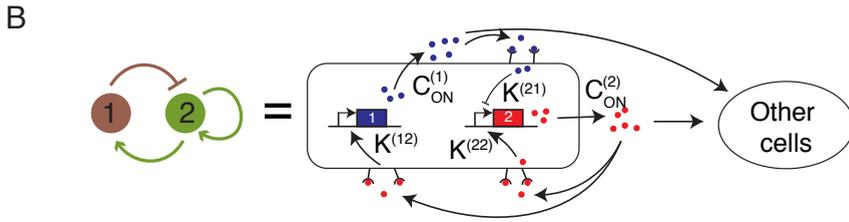
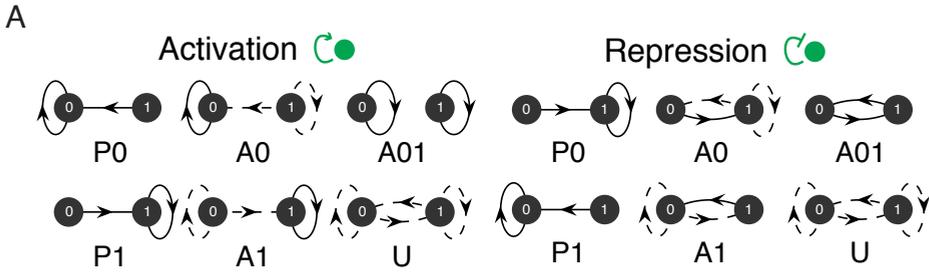
the self-transition. Conversely, not all cell states with a self-transition need to appear in a stationary system state. In other words, having a self-transition does not imply that the state appears in any stationary pattern. As an extreme example, the system could have a fully connected state diagram, where each transition between two cell states is in principle possible. However, this system could still generate a uniform lattice of cells as final state if the parameters are chosen appropriately.

Secondly, the state diagrams show whether periodic steady states (e.g. oscillations) are possible. For any periodic steady state, all cells must revisit their earlier state after  $\tau > 1$  number of time steps, where  $\tau$  is the period of the oscillatory state. This implies that the state diagram should permit cells to return to their initial states after a finite number of time steps, and after passing through other states (otherwise it would be a stationary pattern). This is only possible if the state diagram contains cycles, i.e. closed loops obtained by tracing the edges of the graph from some initial state. The presence of cycles is thus a necessary condition for oscillations. However, it is not a sufficient condition for generating dynamic temporal patterns. This is because it is not guaranteed that a cell can traverse the edges of any cycle one by one when there are possible “routes” on the graph. Each transition then corresponds to a specific condition which depends on the state of all cells of the system. We cannot directly deduce whether a sequence of such transitions is possible at the level of the entire system of  $N$  cells. This is only possible in the special case that all transitions of the cycle are deterministic, i.e. when each node is connected to a unique other node on the graph. We then obtain an oscillation, with a period equal to the length of the cycle.

In summary, the state diagram allows us to deduce two basic properties of our system without running simulations: the set of stationary cell states and the capacity of generating dynamic patterns. These are not purely mathematical properties but have biological relevance. The former is an indicator of multistability and tells us whether a population of identical cells could potentially diversify, generating stable configurations with multiple gene expression profiles. This is known as *phenotypic heterogeneity*, a phenomenon that has seen many experimental studies dedicated to it and may have a wide range of biological functions [Ackermann, 2015]. The latter tells us whether a multicellular system could potentially sustain oscillations, waves or other dynamic patterns (consult the main text for biological examples).

#### SIMPLIFIED DYNAMICS

There are two limits in which the dynamics of the system simplifies dramatically.



1. All interactions are either extremely weak or extremely strong. To be precise, this is the case if each of the interactions is in either the P0 or the P1 phase (all ON/all OFF phase). The system homogenizes after one time step, because each of the interactions is either ON or OFF for all cells in the system. For a spatially uniform system, the dynamics is simple and predictable.
2. All interactions are moderately strong, and the interaction between cells is relatively weak. In more precise terms, suppose all interactions are in the A01 phase. In this case the dynamics of each cell becomes equivalent to that of a single cell. The system is fully autonomous and each cell evolves under its own influence.

In these limits the state diagrams are identical to single-cell state diagrams with rescaled parameters. Therefore, these phases contain only deterministic state diagrams. Thus we know the exact dynamics of the system without running any simulations, for any initial conditions.

5

**Formal derivation** Let us show these two limits more explicitly. Let  $k$  be arbitrary, and  $(i, j)$  be an arbitrary pair of genes with  $M_{\text{int}}^{(ij)} \neq 0$ . Suppose this interaction is in the P0 phase. Then as a result of  $(1 + f_N^{(j)})C_{\text{OFF}}^{(j)} < K^{(ij)}$ , we have

$$g_{P0}^{(ij)} \equiv g_k^{(ij)}(\mathbf{X}) = \begin{cases} 0 & \text{if } M_{\text{int}}^{(ij)} = 1 \\ 1 & \text{if } M_{\text{int}}^{(ij)} = -1 \end{cases}$$

That is to say,  $g_k^{(ij)}(\mathbf{X})$  becomes independent of both  $\mathbf{X}$  and  $k$ . Likewise, in the P1 phase we get  $g_k^{(ij)}(\mathbf{X}) \equiv g_{P1}^{(ij)}$ , with  $g_{P1}^{(ij)} = 1 - g_{P0}^{(ij)}$ . Now suppose all interactions are either in

---

**Figure S5.7 (preceding page): Directed graph (state diagram) representation of reveals all the ways in which a cell's gene expression can change and, in turn, what kinds of patterns can form at the population level (also see Section S5.5.2). (Related to Figure 5.3)** Section S5.5.2 describes in detail how we obtain a directed graph (state diagram) representation for every cellular dialogue. A directed graph representation allows us to determine which cellular dialogue and for which sets of parameter values a periodic steady-state (dynamic patterns) and/or stationary steady-state (static spatial-configurations) can form without having to run any simulations. **(A)** Overview of all state diagrams for one signaling molecule, corresponding to the six different dynamic phases (Section S5.5.2) for self-activation (left) and self-inhibition (right). **(B)** Schematic of a cellular dialogue with two signaling molecules. **(C)** A phase diagram — first introduced in [Maire and Youk, 2015a] and described in Section S5.5.2 — shows the phase (colored region) associated with each regulatory interaction of a cellular dialogue as function of the threshold concentration  $K^{(ij)}$  and maximal secretion rate  $C_{\text{ON}}^{(j)}$  for every regulatory interaction  $i \leftarrow j$ . **(D)** Each phase imposes different constraints on each regulatory interaction of a cellular dialogue. The regulated gene may be activated (or repressed), or the outcome is unknown and depends on other elements (indicated by “?”). **(E-F)** Considering the constraints imposed by the phase diagrams, we constructed state diagrams that summarize all possible temporal changes that a cell's gene-expression can have. Solid lines indicate deterministic transitions, whereas dashed lines indicate that a cell can have multiple possible transitions to a different gene-expression state. **(E)** A state diagram for a population of one cell. **(F)** State diagram for a population of multiple cells.

the P0 or P1 phase. Then,  $X_k^{(j)}(t+1) = \prod_j g_k^{(ij)}(\mathbf{X}) = \prod_j g_{c_j}^{(ij)}$  with  $c_j \in \{P0, P1\}$ . Hence,  $X_k^{(j)}(t+1)$  is also independent of both  $X$  and  $k$ . Therefore, all cells become identical after one time step. For an identical lattice with all cells in a state  $X$ , we note that each cell senses a concentration  $Y^{(j)} = (1 + f_N^{(j)})C^{(j)}(X^{(j)})$ . However,  $K^{(ij)}$  and other parameters are unchanged. Therefore, the evolution of a uniform lattice is equivalent to that of a single cell with a rescaled secretion rate  $C^{(j)}(X) \rightarrow (1 + f_N^{(j)})C^{(j)}(X)$ .

Next, consider the case that all interactions are in the A01 phase. Again, let  $k$  and  $(i, j)$  be arbitrary with  $M_{\text{int}}^{(ij)} \neq 0$ . Let  $\mathbf{X}$  be an arbitrary state of the system, and write  $\mathbf{X} = (X_k, Z_k)$ , with  $Z_k = \{X_l\}_{l \neq k}$ . Then the A01 phase puts the following constraints on the system:

$$\begin{aligned} Y^{(j)}(X_k = 1, Z_k) &\geq C_{\text{ON}}^{(j)} + f_N^{(j)} C_{\text{OFF}}^{(j)} > K^{(ij)}, \\ Y^{(j)}(X_k = 0, Z_k) &\leq C_{\text{OFF}}^{(j)} + f_N^{(j)} C_{\text{ON}}^{(j)} < K^{(ij)}. \end{aligned} \quad (\text{S5.19})$$

We see that any cell with gene  $j$  ON will always satisfy the first constraint, regardless of the rest of the system. Likewise, any cell with gene  $j$  OFF will always satisfy the second constraint. As a result,  $g^{(ij)}(X_k; Z_k) = g^{(ij)}(X_k)$  becomes independent of  $Z_k$ , the states of all cells other than  $k$  in the system. Therefore,  $X_k^{(i)}(t+1) = \prod_j g^{(ij)}(X_k(t); Z_k(t)) = g^{(ij)}(X_k(t))$  depends only on  $X_k(t)$ . In other words, the evolution of any cell in the system is independent of the state of the other cells.

#### ALGORITHM FOR COMPUTING STATE DIAGRAMS

In this subsection, we present a general method for computing the state diagram for a system of one or two genes, given an arbitrary set of system parameters. The construction for two genes can be readily generalized to systems with more than two genes.

For a single gene, we state diagrams follow straightforwardly from the definition of the phases (subsection S5.5.2). The end result can be represented as a directed graph with two nodes (representing ON and OFF state of the gene) and up to four edges, which we can describe using its adjacency matrix

$$A = \begin{pmatrix} A_{00} & A_{01} \\ A_{10} & A_{11} \end{pmatrix}. \quad (\text{S5.20})$$

The adjacency matrix gives information on whether edges are present for each potential link between two nodes. The entries  $A_{ij} \in \{0, 1\}$  are for transitions from state  $i$  to state  $j$ . If  $A_{ij} = 1$ , a transition  $i \rightarrow j$  is possible and we draw an edge between the two nodes. If  $A_{ij} = 0$ ,  $i$  cannot transition to  $j$  and we draw no edge.

As an example, consider cells with a single signaling molecule with negative feedback to

itself. The graphs for positive feedback are deduced in a similar way. In the P1 phase, the system is permanently repressed, so all states go to the 0 state. Hence  $A = \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}$ . By contrast, in the P0 state, both ON and OFF cells always turn ON at the next time step, so  $A = \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix}$ . In the A1 state, ON cells always turn OFF, but we do not know anything about the OFF cells. Hence both transitions  $0 \rightarrow 0$  and  $0 \rightarrow 1$  are possible. Therefore,  $A = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}$ . Conversely, in the A0 phase, only OFF cells are constrained to always turn on, so  $A = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}$ . In the A01 phase, OFF cells turn ON and ON cells turn OFF, so  $A = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ . Finally, in the U phase, all transitions are unconstrained, so  $A = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$ .

## 5

For systems with two or more genes, the procedure of deriving state diagrams is considerably more involved. We first outline the intuitive idea behind this derivation and then provide a formal, mathematical derivation of the construction. With two mutually interacting signaling molecules, the dynamics of a gene  $i$  depends in general on both regulation by itself and regulation by the other gene, which we label  $j$ . If we know the phases of both regulations  $i \leftarrow i$  and  $i \leftarrow j$ , then we can deduce the constraints they impose on the dynamics of  $i$ . To do this, we have to combine the constraints from both regulatory interactions  $i \leftarrow i$  and  $i \leftarrow j$ , for which we employ a three-valued logic operation. Intuitively, this three-valued logic system represents the fact that there are three possible outcomes of each interaction: the regulated gene is activated, the regulated gene is repressed or the outcome is unknown. Hence, we need to know what the final response of gene  $i$  is for each combination of the three outcomes for both of the two regulatory interactions. For instance, suppose that both  $i$  and  $j$  positively regulate  $i$  (i.e.  $M_{\text{int}}^{(ii)} = M_{\text{int}}^{(ij)} = 1$ ), but the interaction  $i \leftarrow i$  is always activating (i.e., the sensed concentration of  $i$  always exceeds the threshold  $K^{(ii)}$ ) while the interaction  $i \leftarrow j$  is unknown. Then the final outcome for gene  $i$  is unknown, because both positive interactions must be activating for the gene to turn on. Next, recall that the phases in general place dynamical constraints that depend on the state of the system. Concretely, this means that the constraint placed by  $i$  on itself depends on whether it is ON or OFF, and the same holds for the constraint placed by gene  $j$ . Therefore, for each combination of states for genes  $i$  and  $j$  — there are four of these, corresponding to the four cell states  $(0, 0)$ ,  $(0, 1)$ ,  $(1, 0)$  and  $(1, 1)$  — we could have a different set of constraints on the dynamics for gene  $i$ . Thus, we have to separately consider each of the four cell states and see which constraints they impose on the dynamics of both gene  $i$  and gene  $j$ . In this way, for each cell state, we obtain all possible cell states to which it can transition to and draw the corresponding edges on the graph. This eventually gives us our state diagram.

**Formal derivation of the algorithm for computing state diagrams** First, we note that the four-node graph with up to 16 edges is now represented by a  $4 \times 4$  adjacency matrix, which we will denote

$$A = \begin{pmatrix} A_{(0,0) \rightarrow (0,0)} & A_{(0,0) \rightarrow (1,0)} & A_{(0,0) \rightarrow (0,1)} & A_{(0,0) \rightarrow (1,1)} \\ A_{(1,0) \rightarrow (0,0)} & A_{(1,0) \rightarrow (1,0)} & A_{(1,0) \rightarrow (0,1)} & A_{(1,0) \rightarrow (1,1)} \\ A_{(0,1) \rightarrow (0,0)} & A_{(0,1) \rightarrow (1,0)} & A_{(0,1) \rightarrow (0,1)} & A_{(0,1) \rightarrow (1,1)} \\ A_{(1,1) \rightarrow (0,0)} & A_{(1,1) \rightarrow (1,0)} & A_{(1,1) \rightarrow (0,1)} & A_{(1,1) \rightarrow (1,1)} \end{pmatrix}. \quad (\text{S5.21})$$

The interpretation is the same: the state  $(i, j)$  can transition into the state  $(k, l)$  if and only if  $A_{(i,j) \rightarrow (k,l)} = 1$ , and we represent this graphically by drawing a directed edge between  $(i, j)$  and  $(k, l)$ . Our goal is then to combine the constraints imposed by the different phases for each interaction to compute this adjacency matrix.

Recall that the time evolution for a cell determined by  $X^{(i)}(t+1) = \prod g^{ij}(\mathbf{X}(t))$ . This is a deterministic equation for  $X^{(i)}(t)$  when we know the precise input system state  $\mathbf{X}(t)$ . Now suppose we only know the cell's own state  $X = (X^{(1)}, X^{(2)})$  and the phase of each interaction  $i \leftarrow j$ . We want to calculate the set of *possible* output cell states for  $X(t+1)$ . To do this, we introduce a set of three-valued logic states  $S = \{0, 1, 2\}$  and a logic AND function  $\wedge : S \rightarrow S$  defined by the truth table

$$\begin{array}{cccc} a \wedge b & b = 0 & b = 1 & b = 2 \\ \begin{array}{l} a = 0 \\ a = 1 \\ a = 2 \end{array} & \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 2 \\ 0 & 2 & 2 \end{pmatrix} & & \end{array} \quad (\text{S5.22})$$

This 3-valued logic system is known as the *Kleene logic*. It has a third logic value UNKNOWN in addition to TRUE and FALSE. In our notation, 0=FALSE, 1=TRUE and 2=UNKNOWN. The UNKNOWN value can be interpreted a state that can be either TRUE or FALSE. When combined with a FALSE value, we know for sure that FALSE  $\wedge$  UNKNOWN = FALSE, since both FALSE  $\wedge$  FALSE = FALSE and FALSE  $\wedge$  TRUE = FALSE. However, TRUE  $\wedge$  FALSE = FALSE while TRUE  $\wedge$  TRUE = TRUE, and therefore TRUE  $\wedge$  UNKNOWN = UNKNOWN.

We employ the three-valued logic system as follows: whenever a cell state has uncertain transitions (i.e. can transition to multiple output states), we assign a value of 2 (UNKNOWN) to it. This also allows us to combine unknown outcomes from different interactions. Any remaining undetermined transitions imply that there are cell states for which multiple transitions are possible.

Concretely, define  $g^{(ij)}(X) \in S$  as the outcome of the interaction  $i \rightarrow j$  for a given input cell states  $X$ . Note that it takes value in  $S$ , indicating that the interaction is either on, off or the outcome is unknown. Let  $Z_{out}(X) = (Z_{out}^{(1)}(X), Z_{out}^{(2)}(X))$  be the three-valued output state given input state  $X$ . We construct the output state as follows:

$$\begin{aligned} Z_{out}^{(1)}(X) &= g^{(11)}(X) \wedge g^{(12)}(X) \\ Z_{out}^{(2)}(X) &= g^{(21)}(X) \wedge g^{(22)}(X). \end{aligned} \quad (S5.23)$$

Here we have replaced ordinary multiplication by the  $\wedge$  operation that takes into account unknown outcomes. This three-valued output state needs to be translated to the actual possible output (binary) cell states of the system. Intuitively, if there is an unknown outcome, i.e.  $Z_{out}^{(i)} = 2$  for some  $i$ , then we should take into account all possible outcomes of that state. Hence we should consider states with both  $Z_{out}^{(i)} = 0$  and  $Z_{out}^{(i)} = 1$ .

Formally, let us denote the set of possible output cell states as  $\Sigma_{out}(X)$ , with elements in  $\{0, 1\}^2$ . We construct the set  $\Sigma_{out}(X)$  through the construction of two maps. Let  $\mathcal{P}(\{0, 1\}) = \{\emptyset, 0, 1, \{0, 1\}\}$  denote the power set of  $\{0, 1\}$ . First we define a map

$$\begin{aligned} \sigma_1 : S &\rightarrow \mathcal{P}(\{0, 1\}) \\ x &\mapsto \begin{cases} x & x \in \{0, 1\} \\ \{0, 1\} & x = 2 \end{cases} \end{aligned} \quad (S5.24)$$

This map constructs the set of possible output gene states, by deconstructing the element  $2 \in S$  into the set  $\{0, 1\}$  of possible outcomes. Extend the map to  $S^2$  by defining  $\sigma : S^2 \rightarrow \mathcal{P}(\{0, 1\})^2$  as  $\sigma(Z) = (\sigma_1(Z^{(1)}), \sigma_1(Z^{(2)}))$ .

Next, we have to put together the deconstructions to arrive at a set of output cell states. Recall that  $\mathcal{X}_1^{(2)} = \{(0, 0), (0, 1), (1, 0), (1, 1)\}$  is the phase space of a single cell with two genes. We define a map

$$\begin{aligned} \tau : \mathcal{P}(\{0, 1\}) \times \mathcal{P}(\{0, 1\}) &\rightarrow \mathcal{P}(\mathcal{X}_1^{(2)}) \\ (x_1, x_2) &\mapsto x_1 \times x_2, \end{aligned} \quad (S5.25)$$

where the  $\times$  denotes an ordinary Cartesian product between sets. For instance, if  $x_1 = \{0, 1\}$  and  $x_2 = 1$ , then  $x_1 \times x_2 = \{(0, 1), (1, 1)\}$ . Hence, the second map constructs all the possible cell states from the possible states for each gene. Putting it together, we construct the set of output cell states as

$$\Sigma_{out}(X) = (\tau \circ \sigma \circ Z_{out})(X). \quad (S5.26)$$

Finally, once we have the output cell states for our input state  $X$ , we set

$$A_{X \rightarrow Y} = 1, \forall Y \in \Sigma_{out}(X). \quad (\text{S5.27})$$

In other words, we draw edges from  $X$  to all cell states in the set of possible output states  $\Sigma_{out}(X)$ . Doing this for all input states  $X$  gives us the full adjacency matrix for the state diagram.

### S5.5.3. ANALYTIC FRAMEWORK FOR TRAVELING WAVE PROPAGATION

In this section, we provide a detailed analysis of traveling waves moving on a constant background of cells, which are found in cellular dialogues 15, 19, 33, 34 and 36. We first discuss features of traveling waves that characterize and distinguish different instances of traveling waves (Section S5.5.3). These features are used in an analytic estimate of the density of traveling wave states in the overall system in Section S5.5.3. The core of this section is composed of a derivation of a set of conditions for TW propagation (Section S5.5.3). We then discuss the performance of the analytic theory in terms of how well it recapitulates simulation results in Section S5.5.3. Finally, we sketch how to extend our method to dynamic patterns on an oscillatory background in Section S5.5.3.

#### FEATURES OF TRAVELING WAVES

The traveling waves that we observe can be distinguished from each other through a number of features:

## 5

1. Orientation and direction of the wave. The waves can be oriented in different ways and for each cellular dialogue we observe waves of all different orientations. We distinguish between horizontally, vertically and diagonally oriented waves (see e.g. Figures 5.2C–E and 5.2G). Horizontal waves wrap around the horizontal axis once, without wrapping around the vertical axis, and travel in the vertical direction. Vertical wave wrap around the vertical axis once, without wrapping around the horizontal axis, and travel in the horizontal direction. Diagonal waves wrap around each of the two axes at least once and can travel in either direction. A more precise way of accounting for the geometry of the wave is through winding numbers, which will be introduced in section S5.5.3.
2. Presence of bends in the wave. We distinguish between straight and bent waves according to whether all cells in a band of the wave are aligned in the same direction. For a triangular lattice, there are three directions along which the cells can align themselves. In one case, we get straight horizontal waves (e.g. Figure 5.2C), whereas in the other cases we get diagonally oriented waves (e.g. Figure 5.2G). However, we can also get waves with one or more bends (e.g. Figure 5.2E), points at which the alignment of the cells changes direction. Note that the cells located at the bends have a different set of nearest neighbors from the aligned cells. Furthermore, we can distinguish between bends that are in the direction of propagation (outward bends) and bends that are opposite to the direction of propagation (inward bends).
3. Number of waves. In the simplest case, the system self-organizes into a single wave on a uniform background (e.g. Figures 5.2C and 5.2E). However, we also ob-

serve multiple coexisting waves, separated by each other by regions of cells with the background state (see e.g. Figures 5.2A and 5.2D). Such waves have the same orientation and direction of travel, but are not necessarily aligned parallel with each other.

4. Number of different cell states in the wave. For almost all the waves we observed, we found wave made up of three different cell states. The background was made up of the fourth cell states. The exact states which make up the wave and their order varies from topology to topology, and sometimes also between different parameter sets of the same topology. In rare cases, we also found waves consisting of two types of cell states on a background of a third cell state.
5. Number of bands in the wave. In most cases, we find waves consisting of single bands of cells of the same state. Waves with bands with two or more layers of cells and waves where different cell states have different band widths have also been observed (see e.g. Figure 5.2C).
6. Defects. In rare cases we may see waves which contain single-cell defects such as an additional cell of the same cell state attached to an otherwise normal wave.

#### ABUNDANCE OF TRAVELING WAVES

We now derive an estimate of the relative abundance of traveling waves of the forms we observe in the system. Due to the variety of morphologies these waves can take, we could expect them to take up a considerable portion of the total phase space. In this scenario, finding system conditions under which most of the simulations go to traveling waves would not be entirely surprising. On the other hand, if the relative abundance of traveling waves in the system is low, we could interpret this as a sign that there is a self-organizing mechanism that drives the system towards traveling wave formation.

First, we identify the key aspects of traveling waves and divide them into a limited number of categories. For each category, we then calculate the number of distinct shapes the waves can take, as well as the total number of distinct “snapshots” each wave form is made up of. This then gives us an estimate of the total number of states in the system that can be considered traveling waves.

We have previously provided a list of features that distinguish traveling waves from each other. While we have observed waves that differ in all these categories, we note that the vast majority of waves have the same features for a number of categories. In particular, most waves are composed of three cell states (with a fourth background state), are com-

posed of a single band and have no defects. We also rarely observe more than two waves propagating simultaneously in moderately large systems (e.g.  $N = 256$ ). Hence we only consider the orientation and direction, the presence of bends and the number of waves to account for the vast majority of observed wave forms. In the following we consider a generic single-banded wave with  $N = n^2$  cells.

The orientation of a wave can be made more precise by considering the number of times the wave wraps around each axis. Since the system is periodic, effectively we are dealing with wave that winds around each axis of a torus different numbers of times. Let  $W_x, W_y$  be the winding numbers around the horizontal and vertical axis. For a plane horizontal wave such as shown in Figure 5.2A,  $W_x = 1, W_y = 0$ . For a vertical wave such as in Figure 5.2E,  $W_x = 0, W_y = 1$ . The diagonal wave in Figure 5.2G has  $W_x = 1, W_y = 2$ , but we can also imagine diagonal waves that wrap around the system in different ways. The most common winding numbers are listed in Table S5.2. As is apparent from the table, we mostly observe simple waves that are either horizontal, vertical or diagonal, but wrap around the axes only a few times. Note that traveling waves are characterized by  $W_x + W_y \geq 1$ , i.e. a traveling wave always wraps at least once around one of the axes. (Smaller structures that do not wrap around either axis but do translate in space are referred to as traveling pulses).

5

$W_x$	$W_y$	$N_C(W_x, W_y)$	$N_{wf}(W_x, W_y)$	$T(W_x, W_y)$
1	0	$n$	1	$n$
0	1	$n$	$\begin{pmatrix} n \\ \frac{1}{2}n \end{pmatrix}$	$n$
1	1	$\frac{3}{2}n$	$\begin{pmatrix} \frac{3}{2}n \\ \frac{1}{2}n \end{pmatrix}$	$2n$
1	2	$2n$	1	$2n$
2	1	$\frac{5}{2}n$	$\begin{pmatrix} \frac{5}{2}n \\ n \end{pmatrix}$	$2n$

**Table S5.2:** Main properties of most common types of waves.  $n$  is the linear grid size, with  $N = n^2$ . We assume that  $n$  is an even number, so that the system is a perfect hexagonal lattice on a torus. The data is based on empirical observations of self-generated traveling waves.  $W_x, W_y$  are the winding numbers,  $N_C$  is the number of cells of the wave,  $N_{wf}$  is the number of wave forms and  $T$  is the period of the wave.

Once we fixed the winding numbers, the precise form of the wave is often still unspecified. For instance, a vertical wave can have different number of bends in both directions. Nevertheless, we can derive a general formula for the number of wave forms given  $(W_x, W_y)$ . Let us consider a single wave that travels in a fixed direction. Suppose we pick a random cell of the wave. Empirically, we find that each cell of the wave that has the same state has precisely two neighbors with the same state. This is even the case when

there are complicated bends in the wave. Now pick one of the neighbors of our selected cell that has the same cell state. The nearest-neighbor vector that connects the two cells lies along one of the six directions one can travel in on a hexagonal lattice. These can be expressed in terms of the basis vectors of the lattice as  $\vec{e}_1, -\vec{e}_1, \vec{e}_2, -\vec{e}_2, \vec{e}_3 \equiv \vec{e}_2 - \vec{e}_1, -\vec{e}_3 = \vec{e}_1 - \vec{e}_2$  (Figure S5.8B – left figure). The second cell has a unique neighbor of the same state that we have not selected yet. The vector between the second and third cell defines a new direction that we record. We can therefore continue this procedure and pick subsequent cells in our wave, until we get back to our original cell. This is because the wave wraps around an axis at least once as noted before. For each step we take, we keep track of the direction we need to move in to get to the next cell. At the end, we count the number of steps in each of the six directions obtained through this procedure (illustrated in Figure S5.9B – right figure). Let us denote these by  $\{n_{i,\alpha}\}$ , where  $1 \leq i \leq 3$  and  $\alpha \in \{-, +\}$ . For example,  $n_{2,-}$  gives the number of steps we took in the  $-\vec{e}_2$  direction.

Wave forms differ by their set of nearest-neighbor vectors that we obtain with this procedure. Nevertheless, once we fix the winding numbers, this constraints the possible sets of direction vectors in a way we can make precise. First, we note that empirically we find that waves with fixed winding numbers always have the same number of cells of a given state, which we will denote  $N_C(W_x, W_y)$ . Empirical results for commonly found waves are listed in Table S5.2. For instance, for a horizontal wave, we find that it always has  $N_C(1, 0) = n$  cells of a given state, which make up exactly one row of the lattice. This constrains the total number of nearest-neighbor vectors to  $N_C(W_x, W_y)$ , such that our first constraint is

$$\sum_{i,\alpha} n_{i,\alpha} = n_{1,-} + n_{1,+} + n_{2,-} + n_{2,+} + n_{3,-} + n_{3,+} = N_C(W_x, W_y) \quad (\text{S5.28})$$

Next, the winding numbers constrain the number of occurrences of each nearest-neighbor vector. For instance, for a horizontal wave the nearest neighbor vectors when added up must be align in the horizontal direction, with a magnitude equal to the grid size. However, a priori this does not imply that all nearest neighbor vectors are in the  $\vec{e}_1$  direction, since  $\vec{e}_2$  and  $\vec{e}_3$  also have horizontal components. In general, the constraints are that the number of steps taken in the horizontal and vertical directions must be equal to the  $\pm W_x n$  and  $\pm W_y n$  in order to return to the original cell. The sign degeneracy comes from the fact that starting from the initial cell we pick, we can traverse the wave in two different directions, which yield winding numbers that differ by a minus sign. Working out

these conditions, we derive the following constraints:

$$n_{1,-} - n_{1,+} - \frac{n_{2,-}}{2} + \frac{n_{2,+}}{2} + \frac{n_{3,-}}{2} - \frac{n_{3,+}}{2} = \pm W_x n, \quad (\text{S5.29})$$

$$-n_{2,-} + n_{2,+} - n_{2,+} + n_{3,+} = \pm W_y n \quad (\text{S5.30})$$

We can now try to solve these constraints together with the general constraints  $0 \leq n_{i,\alpha} \leq n$  for all  $i, \alpha$  for given winding numbers  $W_x, W_y$ . For all the winding numbers listed in Table S5.2, we obtained solutions of the form  $n_{i_1,\alpha_1} = n_{i_1,\alpha_1}(n) > 0$ ,  $n_{i_2,\alpha_2} = n_{i_2,\alpha_2}(n) > 0$  for some  $i_1, \alpha_1$  and  $i_2, \alpha_2$ ,  $n_{i,\alpha} = 0$  for all other  $i, \alpha$ . The interpretation of this result is that in practice all waves are formed by traveling continuously in two directions, i.e. they never “bend back”. Secondly, we found that the number of steps in each direction is a linear function  $n$ , so we get an explicit scaling of our results with system size.

## 5

We can now readily obtain the number of wave forms that satisfy these constraints. This reduces to a simple combinatorics problem where the wave forms differ by the order in which the directions  $i_1, \alpha_1$  and  $i_2, \alpha_2$  appear, in the procedure described above. Since there are  $N_C(W_x, W_y)$  such vectors (one for each cell of a wave layer), there are  $N_{wf}(W_x, W_y) \equiv \binom{N_C(W_x, W_y)}{n_{i_1, \alpha_1}} = \binom{N_C(W_x, W_y)}{n_{i_2, \alpha_2}}$  ways of ordering the vectors in either direction, corresponding to the number of possible wave forms with the given winding numbers. Finally, the sign degeneracy of  $W_x, W_y$  introduces an additional factor of 2 whenever both  $W_x, W_y > 0$ . There are the four possibilities  $(W_x, W_y)$ ,  $(W_x, -W_y)$ ,  $(-W_x, W_y)$ , and  $(-W_x, -W_y)$ , but only  $(W_x, W_y)$  and  $(-W_x, -W_y)$  give equivalent waves, whereas  $(-W_x, W_y) \equiv (W_x, -W_y)$  yields a different wave, corresponding to a different diagonal orientation.

The direction of a wave can in principle be in any of the six directions the hexagonal lattice allows. However, once the orientation of a wave is fixed, there are only two possible directions remaining. For instance, for a horizontal wave, the only directions are up and down.

The number of waves that can simultaneously propagate depends on the system size. For  $N = 256$ , we rarely observe more than two simultaneously propagating waves. Note that the waves need to have the same orientation and direction of motion, or else they would collide and annihilate or form new waves. Once the shapes of both waves are fixed, an additional variable is the spacing between the waves. Assume that both waves are horizontal, then the variable is the number of rows between the waves. For two

waves, the distance between the waves lies in the range  $[1, \frac{n-6}{2}]$ , with  $\frac{n-6}{2}$  being an upper bound for when both waves are straight. This gives a degeneracy of roughly  $\frac{n-6}{2}$ . This is because both waves take up 3 rows, and from the arrangement of the remaining rows we take the shortest distance since the system is periodic.

Putting everything together, we now obtain our general estimate for the density of traveling waves in phase space. We estimate this to be in the order of

$$N n_{dir} \sum_{n_{waves}} n_{dist}(n_{waves}) \sum_{W_x, W_y} D(W_x, W_y) N_{wf}(W_x, W_y)^{n_{waves}}. \quad (S5.31)$$

Here  $n_{dir} = 2$  signifies the two directions of propagation, the first sum is over the total number of waves in the system  $n_{waves}$ , the number of unique distances between the waves is denoted  $n_{dist}$ , the degeneracy after accounting for negative winding numbers is denoted  $D(W_x, W_y)$ , and the second sum is over the winding numbers  $W_x, W_y \in \mathbb{N}_0$ . From the previous part we estimate  $n_{dist}(1) = 1$ ,  $n_{dist}(2) = \frac{n-6}{2}$ . The final term signifies the fact that in case of multiple waves, they can in principle have independent wave forms (with the same direction and winding numbers). The term  $N$  in front accounts for the possible positions of the wave on the lattice, obtained simply by counting the number of ways to place a given cell of the wave on the lattice. This is an upper bound since in case waves with symmetry different placements of this selected cell could still give the same configuration.

**Traveling wave formation time** The aim of doing the traveling wave density estimation is to provide more direct evidence for the intuitive idea that a self-organizing mechanism drives the formation of traveling waves (TWs), and that these do not simply arise by chance because TW states are abundant. To do so, we looked at the TW formation time, i.e. the number of time steps it takes to go from a random initial state to a TW. We compared empirically observed formation times from simulations with expected simulation times for a random process based on our wave density estimation. The random process can be seen as a null hypothesis stating that there is no self-organization, but rather TWs form by chance as the system randomly samples states in the system. To show that there is a self-organization process, we therefore demonstrate that the findings on TW formation times differ considerably from those of a random process. More precisely, we considered a stochastic process whereby subsequent states of the system would be randomly sampled over the set of all  $2^N$  states with a uniform distribution, i.e. each next state would be drawn from the set of all states with equal probability for each of the  $2^N$  states. Under this assumption, the formation time of a TW would be

equivalent to the first time of success in a Bernoulli process where the probability of success  $p$  equals the density of TWs derived above (Equation S5.31). Hence the waiting time — here the TW formation time — would follow a geometric distribution with the same parameter  $p$ , with an expected waiting time of  $1/p$ .

We first considered the expected waiting time as a function of the system size (measured by the grid size  $\sqrt{N}$ ). For the random process described above, we obtained a mean formation time that scales faster than exponentially with the grid size (Figure S5.8C – left plot), and quickly reaches times that are orders of magnitude larger in our simulations. Conversely, our simulated data shows that the mean formation time scales linearly with grid size (Figure S5.8C – right plot), and has typical values well within reasonable simulation time bounds.

Given this result, it is still possible that our simulations probe only the waves that form within the limited simulation time, while the majority of other TWs have formation times that are orders of magnitude larger. However, in this case we would obtain large fractions of simulations that do not reach any steady state at our maximum simulation time, which is not the case (see for instance Figure S5.12A — here the fraction of simulations reaching  $t_{\max}$  is low across most parameter sets capable of generating TWs). Furthermore, a second counterargument is that the TW formation time distribution would then be skewed towards the right, with relatively more TWs taking times closer to the maximum simulation time. This would be necessary to produce the geometric distribution of the Bernoulli process (Figure S5.8D – left plot). In contrast, we observed the opposite — the simulated formation time distribution is skewed toward the left, with relatively more wave forming within a short amount of time (Figure S5.8D – right plot).

In conclusion, we have provided evidence that traveling waves form through a self-organizing process whereby the system ‘actively’ converges onto traveling wave attractor states, rather than ‘passively’ exploring the state space of all possible configurations and randomly finding traveling wave attractors. This result should be intuitive given the relative ease of finding traveling waves in our simulations. Note that we have not aimed to give a rigorous definition of self-organization in this context or formally proving our claims, but merely want to point out that the self-organization process differs significantly from a purely random process.

#### TRAVELING WAVE PROPAGATION CONDITIONS

**General conditions for pattern propagation** Before working out the case of traveling waves in detail, let us discuss what pattern propagation means in general. Suppose we have a pattern that periodically repeats itself in time. All information about the pattern is encoded in the states of the pattern over one period. Let us denote these by  $\{X(0), X(1), \dots, X(\tau)\}$ , where  $X(t) = \{X_k^{(i)}(t)\}_{1 \leq i \leq l, 1 \leq k \leq N}$  is the state of the system at time  $t$  as

specified by the states of each of the genes of each cell. These can be considered a series of snapshots of the system that make up a movie of the dynamic pattern when played. In general, take an arbitrary state  $\xi(t)$  and suppose that the system is updated according to a rule

$$\xi(t+1) = f(\xi(t); P), \quad (\text{S5.32})$$

for some unspecified function  $f$  of the current state of the system that depends on the parameters of the system denoted by  $P$ . The condition that the pattern can propagate under the set of parameters  $P$  is precisely that  $f$  updates each snapshot of the pattern to the next snapshot of the system. In other words,  $X(t+1) = f(X(t); P)$  for all  $0 \leq t \leq \tau - 1$ .

In general, this would put constraints on each of the cells of the system, leading to a convoluted set of conditions for pattern propagation. However, in cases where the pattern exhibits a symmetry, these conditions can be drastically simplified. In one extreme case, if the pattern is a homogeneous collection of identical cells, at any time we would only need to check one set of conditions for an arbitrary cell of the system. Conversely, suppose the system is completely anisotropic for the whole duration of the pattern trajectory. Each cell then sees a different environment at any time. We then need to check all  $N \times l \times \tau$  conditions for each of the genes of each of the cells of the system, at each time step of the system.

The case of traveling waves allows us to exploit the symmetry of the pattern to drastically reduce the number of conditions for pattern propagation. First note that traveling waves are characterized by the fact that the state upon updating is related to the previous state by a simple translation in space. This means that rather than checking conditions for each time step, we only need to check that the wave propagates at one arbitrary time step. Furthermore, the spatial symmetry of the system allows us to check only a small number of cells of the system, as will be explained in the next section.

**Straight and bent waves** To derive conditions for the propagation of these waves, we look at straight (plane) waves (Figure 5.4A, “straight wave”), waves with a single outward bend (Figure 5.4A, “bent wave”) and waves with a single inward bend (Figure 5.4A, “bent wave” with reverse direction of propagation). In this way, we obtain results applicable to the vast majority of waves observed in the system, including more complicated waves which are typically locally still similar to these “simple” waves. For instance, the configuration in Figure 5.1F contains two waves with multiple bends. However, the nearest neighbors of any cell is identical to the neighbor structure of a cell in one of the three

prototype waves. Namely, the cells at the tip of the wavefront have nearest neighbors that is identical to the cell at the tip of the outward bent single wave. The cells that are bent towards the back of the wavefront have a neighbor that is identical to those at the bend of the inward bent wave.

Therefore, it suffices to study the conditions for propagation of each of these three simple types of waves. This gives a first approximation to the propagation conditions for more complicated waves, and is valid especially when the interaction between cells is not too strong and takes place mostly on a local scale (i.e. when  $a_0$  is sufficiently high). The types of waves which are not covered by this analysis are waves with multiple bands (because the cells of such waves have different nearest neighbors), and waves with defects (which are too rare to motivate analysis of each special case).

## 5

**Structure of traveling waves** All waves consisting of three consecutive single bands of cells have a similar spatial structure. For any instance of such a wave, we can identify six types of cells that each have a unique set of nearest neighbors. Let us denote these six types of cells as follows (see Figure 5.4B):

1.  $E_F$  – front exterior
2.  $F$  – front
3.  $M$  – middle
4.  $B$  – back
5.  $E_B$  – back exterior
6.  $E$  – exterior

Note that the types  $E$ ,  $E_F$  and  $E_B$  all have the same cell state (the state of the white color in Figure 5.2A). However, we divide the exterior cells up into three classes because they have different sets of nearest neighbors. A cell of type  $E_F$  in front of the wave neighbors  $F$  cells, whereas a cell  $E_F$  at the back of the wave neighbors  $B$  cells, while the rest of the  $E$  cells border only other  $E$  cells.

Hence, the six types of cells have four different cell states, which we denote as  $X(F)$ ,  $X(M)$ ,  $X(B)$  and  $X(E) = X(E_F) = X(E_B)$ . For binary cells, the possible cell states form the set  $S = \{(0, 0), (0, 1), (1, 0), (1, 1)\}$ .

<b>Plane wave</b>				
Wave state	Number of neighbor states			
	$X(F)$	$X(M)$	$X(B)$	$X(E)$
$E_F$	2	0	0	4
$F$	2	2	0	2
$M$	2	2	2	0
$B$	0	2	2	2
$E_B$	0	0	2	4
$E$	0	0	0	6

<b>Wave with outward bend</b>				
Wave state	Number of neighbor states			
	$X(F)$	$X(M)$	$X(B)$	$X(E)$
$E_F$	1	0	0	5
$F$	2	1	0	3
$M$	3	2	1	0
$B$	0	3	2	1
$E_B$	0	0	3	3
$E$	0	0	0	6

<b>Wave with inward bend</b>				
Wave state	Number of neighbor states			
	$X(F)$	$X(M)$	$X(B)$	$X(E)$
$E_F$	3	0	0	3
$F$	2	3	0	1
$M$	1	2	3	0
$B$	0	1	2	3
$E_B$	0	0	1	5
$E$	0	0	0	6

**Table S5.3:** Cell states of nearest neighbors of the six types of cells ( $E_F, F, M, B, E_B, E$ ) for straight waves and for the cells at the tip of waves with bends (Figure 5.4B). Results are for a hexagonal lattice with coordination number  $z = 6$ .

At any straight segment of a wave, the cells of the wave and those bordering the wave have exactly the same local structure (nearest neighbors). Concretely, this means that any cell of the straight segment borders the same number of cells of each state (Table 1). For instance, an  $F$  cell will always border two cells with state  $X(F)$ , two cells with state  $X(M)$  and two cells with state  $X(E)$ .

The set of nearest neighbors of a bent wave differs from that of plane waves only at the location of the bend. The rest of the cells have nearest neighbors identical to plane wave cells (Figure 5.4B). We therefore take the propagation condition for the cells at the bend into account separately (Table S5.3).

**Traveling wave propagation conditions** For a traveling wave to propagate, we need a number of conditions to be satisfied. Since traveling waves have the property that the entire structure translates forward by one step, we can easily find these conditions. Basically, all cells of one layer take up the state of the next layer and the background cells remain constant. For example, an  $E_F$  cell right in front of the wave should become an  $F$  cell at the next time step. Hence we require that the cell obtains the state  $X(F)$  upon updating. Let  $\alpha \rightarrow X(\alpha')$  denote the condition that a cell of type  $\alpha$  acquires state  $X(\alpha')$  according to the update rule. Then we can succinctly write our set of conditions as:

$$\begin{aligned}
 E_F &\rightarrow X(F) \\
 F &\rightarrow X(M) \\
 M &\rightarrow X(B) \\
 B &\rightarrow X(E) \\
 E_B &\rightarrow X(E) \\
 E &\rightarrow X(E)
 \end{aligned} \tag{S5.33}$$

For a straight wave without bends, these conditions need to be checked only once, for cells that have the nearest neighbors as detailed in the first table in Table S5.3. For waves with at least one bend, both the condition at the location of the bend (either outward or inward) as well as the condition for plane waves (for the straight segments of the wave) need to be checked. For waves with a zig-zag pattern that have no straight segments (e.g. Figure 5.2D), only the conditions for inward and outward bends need to be checked.

Mathematically, we can represent the propagation conditions through a general set of equations in terms of the network topology (specified by  $M_{\text{int}}^{(ij)}$ ), the sensed concentration ( $Y_\alpha^{(j)}$  for a cell of type  $\alpha$ ), and the sensing threshold  $K^{(ij)}$ , which for AND-logic signal integration takes the following form:

$$X^{(i)}(\alpha) = \prod_j \left[ \theta \left( (Y_\alpha^{(j)} - K^{(ij)}) M^{(ij)} \right) + 1 - \left| M^{(ij)} \right| \right] \tag{S5.34}$$

Here  $X^{(i)}(\alpha)$  represents the state of gene  $i$  the cell with state  $\alpha$  should transition into. Here we define  $\theta(x)$  as the Heaviside function with convention  $\theta(0) = 0$ , i.e.  $\theta(x) = \begin{cases} 0 & x \leq 0 \\ 1 & x > 0 \end{cases}$ . Likewise, for OR-logic we have (see Equation S5.11):

$$X^{(i)}(\alpha) = \sum_j \theta \left( (Y_\alpha^{(j)} - K^{(ij)}) M^{(ij)} \right) - \prod_j \theta \left( (Y_\alpha^{(j)} - K^{(ij)}) M^{(ij)} \right). \tag{S5.35}$$

**Nearest-neighbor approximation** Given an exact form of the wave and a specific interaction network of the two genes, we can work out the six conditions for traveling wave propagation, to obtain exact conditions in terms of system parameters. However, since the waves can have different features, we look for a more general approach that predicts propagation independent of the precise shape of the wave. To do this, we will apply a nearest-neighbor approximation (NNA). The idea is to only consider the immediate neighbors of a cell when calculating the concentration it senses, and take into account the rest of the cells through averaging and assuming they are randomly distributed.

Write  $Y_\alpha^{(i)}$  for the concentration of molecule  $i$  a cell of type  $\alpha$  senses. Then we can split the sensed concentration into terms of the cell itself, its neighbors and an approximation of the rest of the lattice, which we assume to be independent of the cell type  $\alpha$ :

$$Y_\alpha^{(i)} = Y_{self}^{(i)}(\alpha) + Y_{nei}^{(i)}(\alpha) + Y_{MF}^{(i)}. \quad (S5.36)$$

Recall that  $C^{(i)}(X)$  is the secretion rate for molecule  $i$ . Denote  $f_{nn}^{(i)} = f^{(i)}(a_0)$  as the nearest neighbor interaction strength, and  $n(X; \alpha)$  as the number of cells of state  $X$  that neighbor a cell of type  $\alpha$ . The sensed concentration due to neighbors can then be written as

$$Y_{nei}^{(i)}(\alpha) = \sum_{X \in S} f_{nn}^{(i)} n(X; \alpha) C^{(i)}(X). \quad (S5.37)$$

The contribution of rest of the lattice is estimated through a mean-field approximation. For a wave with  $N_w$  waves, each consisting of bands of width  $W$ , with winding numbers  $W_x, W_y$  (Section S5.5.3), we can calculate the proportion of cells that have either of the genes on. This proportion depends on the cell states of the wave and background cells. In general, the fraction of cells with a given gene on is

$$p^{(i)} = \frac{N_w W}{N_C(W_x, W_y)} \left( X^{(F)(i)} + X^{(M)(i)} + X^{(B)(i)} \right) + \left( 1 - 3 \frac{N_w W}{N_C(W_x, W_y)} \right) X^{(E)(i)} \quad (S5.38)$$

Here  $X(S)^{(i)}$  denotes the state of gene  $i$  of a cell state  $X(S)$ ,  $N_C(W_x, W_y)$  is the number of cells of one layer given winding numbers  $(W_x, W_y)$  as used in the wave density estimation (Section S5.5.3). The mean-field contribution is then estimated to be the interaction strength of all cells excluding the nearest neighbors times the average secretion rate of the cells:

$$Y_{MF}^{(i)} = \left( f_N^{(i)} - 6f_{nn}^{(i)} \right) \left[ C_{ON}^{(i)} p^{(i)} + C_{OFF}^{(i)} \left( 1 - p^{(i)} \right) \right]. \quad (S5.39)$$

In practice, we looked at single waves ( $N_w = 1$ ) with layers of a single cell thick ( $W = 1$ ), and only considered horizontal and vertical waves, for which  $N_C(W_x, W_y) = n = \sqrt{N}$

(Table S5.2). In this case, the expression for  $p^{(i)}$  simplifies to:

$$p^{(i)} = \frac{1}{\sqrt{N}} \left( X(F)^{(i)} + X(M)^{(i)} + X(B)^{(i)} \right) + \left( 1 - \frac{3}{\sqrt{N}} \right) X(E)^{(i)}. \quad (\text{S5.40})$$

#### EXPLICIT EXAMPLE

In this section, we work out an explicit example of the propagation conditions we derived. Consider cellular dialogue 15, which has interaction matrix  $M_{\text{int}} = \begin{pmatrix} 1 & -1 \\ 1 & 0 \end{pmatrix}$ . From simulations we observed traveling waves with the composition

- $X(F) = (1, 0)$ ,
- $X(M) = (1, 1)$ ,
- $X(B) = (0, 1)$ ,
- $X(E) = (0, 0)$ .

**Conditions for propagation** Let us denote  $E_F = (0, 0)_F$  and  $E_B = (0, 0)_B$ . Explicitly, the conditions for propagation can be expressed in terms of inequalities (Equation S5.41).

Condition	Condition on gene 1	Condition on gene 2
$(0, 0)_F \rightarrow (1, 0)$	$Y_{(0,0)_F}^{(2)} < K^{(12)} \wedge Y_{(0,0)_F}^{(1)} > K^{(11)}$	$Y_{(0,0)_F}^{(1)} < K^{(21)}$
$(1, 0) \rightarrow (1, 1)$	$Y_{(1,0)}^{(2)} < K^{(12)} \wedge Y_{(1,0)}^{(1)} > K^{(11)}$	$Y_{(1,0)}^{(1)} > K^{(21)}$
$(1, 1) \rightarrow (0, 1)$	$Y_{(1,1)}^{(2)} > K^{(12)} \vee Y_{(1,1)}^{(1)} < K^{(11)}$	$Y_{(1,1)}^{(1)} > K^{(21)}$
$(0, 1) \rightarrow (0, 0)$	$Y_{(0,1)}^{(2)} > K^{(12)} \vee Y_{(0,1)}^{(1)} < K^{(11)}$	$Y_{(0,1)}^{(1)} < K^{(21)}$
$(0, 0)_B \rightarrow (0, 0)$	$Y_{(0,0)_B}^{(2)} > K^{(12)} \vee Y_{(0,0)_B}^{(1)} < K^{(11)}$	$Y_{(0,0)_B}^{(1)} < K^{(21)}$
$(0, 0)_R \rightarrow (0, 0)$	$Y_{(0,0)_R}^{(2)} > K^{(12)} \vee Y_{(0,0)_R}^{(1)} < K^{(11)}$	$Y_{(0,0)_R}^{(1)} < K^{(21)}$

(S5.41)

We can show that the last condition for  $(0, 0)_R$  is redundant in general. Namely, we have  $Y_{(0,0)_R}^{(i)} \leq Y_{(0,0)_F}^{(i)}$  and  $Y_{(0,0)_R}^{(i)} \leq Y_{(0,0)_B}^{(i)}$  for both molecules  $i = 1, 2$ . This implies that if the condition for gene 2 for  $(0, 0)_B$  is fulfilled, then the condition for  $(0, 0)_R$  is automatically fulfilled, since  $Y_{(0,0)_R}^{(i)} \leq Y_{(0,0)_B}^{(i)} < K^{(21)}$ . For gene 1, the conditions  $Y_{(0,0)_F}^{(2)} < K^{(12)}$  (condition on  $(0, 0)_F$ ) and  $Y_{(0,0)_R}^{(2)} > K^{(12)}$  (condition on  $(0, 0)_R$ ) give a contradiction. Since the first condition has to be true due to the AND function, the second is necessarily false. This leaves  $Y_{(0,0)_R}^{(1)} < K^{(11)}$ . But since  $Y_{(0,0)_R}^{(1)} = Y_{(0,0)_B}^{(1)}$ , this becomes equivalent to the condition  $Y_{(0,0)_B}^{(1)} < K^{(11)}$  (condition on  $(0, 0)_B$ ). Furthermore, we also have  $Y_{(0,0)_B}^{(2)} = Y_{(1,0)}^{(2)}$ , and since  $Y_{(1,0)}^{(2)} < K^{(12)}$ , the condition  $Y_{(0,0)_B}^{(1)} > K^{(12)}$  must be false. Thus,  $Y_{(0,0)_B}^{(1)} < K^{(11)}$  be-

comes the only condition for gene 1 for  $(0,0)_B$ . Note that these arguments are specific to the network and wave form under consideration and do not need to hold in general.

**Plane waves** We work out the equations explicitly for plane waves. From Equation S5.36-S5.39 and Table S5.3, we get the sensed concentrations

$$\begin{aligned}
 Y_{(0,0)_F}^{(1)} &= 1 + 2C_{ON}^{(1)}f_{nn}^{(1)} + 4f_{nn}^{(1)}, & Y_{(0,0)_F}^{(2)} &= 1 + 6f_{nn}^{(2)}, \\
 Y_{(1,0)}^{(1)} &= C_{ON}^{(1)} + 4C_{ON}^{(1)}f_{nn}^{(1)} + 2f_{nn}^{(1)}, & Y_{(1,0)}^{(2)} &= 1 + 2C_{ON}^{(2)}f_{nn}^{(2)} + 4f_{nn}^{(2)}, \\
 Y_{(1,1)}^{(1)} &= C_{ON}^{(1)} + 4C_{ON}^{(1)}f_{nn}^{(1)} + 2f_{nn}^{(1)}, & Y_{(1,1)}^{(2)} &= C_{ON}^{(2)} + 4C_{ON}^{(2)}f_{nn}^{(2)} + 2f_{nn}^{(2)}, \\
 Y_{(0,1)}^{(1)} &= 1 + 2C_{ON}^{(1)}f_{nn}^{(1)} + 4f_{nn}^{(1)}, & Y_{(0,1)}^{(2)} &= C_{ON}^{(2)} + 4C_{ON}^{(2)}f_{nn}^{(2)} + 2f_{nn}^{(2)}, \\
 Y_{(0,0)_B}^{(1)} &= 1 + 6f_{nn}^{(1)}, & Y_{(0,0)_B}^{(2)} &= 1 + 2C_{ON}^{(2)}f_{nn}^{(2)} + 4f_{nn}^{(2)},
 \end{aligned} \tag{S5.42}$$

We can then write out the conditions from Equation S5.41 explicitly. For gene 1 this gives:

$$\begin{aligned}
 (0,0)_F \rightarrow (1,0): & \quad 1 + 6f_{nn}^{(2)} + Y_{MF}^{(2)} < K^{(12)} \wedge 1 + 2C_{ON}^{(1)}f_{nn}^{(1)} + 4f_{nn}^{(1)} + Y_{MF}^{(1)} > K^{(11)} \\
 (1,0) \rightarrow (1,1): & \quad 1 + 2C_{ON}^{(2)}f_{nn}^{(2)} + 4f_{nn}^{(2)} + Y_{MF}^{(2)} < K^{(12)} \wedge \\
 & \quad C_{ON}^{(1)} + 4C_{ON}^{(1)}f_{nn}^{(1)} + 2f_{nn}^{(1)} + Y_{MF}^{(1)} > K^{(11)} \\
 (1,1) \rightarrow (0,1): & \quad C_{ON}^{(2)} + 4C_{ON}^{(2)}f_{nn}^{(2)} + 2f_{nn}^{(2)} + Y_{MF}^{(2)} > K^{(12)} \vee \\
 & \quad C_{ON}^{(1)} + 4C_{ON}^{(1)}f_{nn}^{(1)} + 2f_{nn}^{(1)} + Y_{MF}^{(1)} < K^{(11)} \\
 (0,1) \rightarrow (0,0): & \quad C_{ON}^{(2)} + 4C_{ON}^{(2)}f_{nn}^{(2)} + 2f_{nn}^{(2)} + Y_{MF}^{(2)} > K^{(12)} \vee \\
 & \quad 1 + 2C_{ON}^{(1)}f_{nn}^{(1)} + 4f_{nn}^{(1)} + Y_{MF}^{(1)} < K^{(11)} \\
 (0,0)_B \rightarrow (0,0): & \quad 1 + 6f_{nn}^{(1)} + Y_{MF}^{(1)} < K^{(11)}
 \end{aligned} \tag{S5.43}$$

For gene 2, the conditions are

$$\begin{aligned}
 (0,0)_F \rightarrow (1,0): & \quad 1 + 2C_{ON}^{(1)}f_{nn}^{(1)} + 4f_{nn}^{(1)} + Y_{MF}^{(1)} < K^{(21)} \\
 (1,0) \rightarrow (1,1): & \quad C_{ON}^{(1)} + 4C_{ON}^{(1)}f_{nn}^{(1)} + 2f_{nn}^{(1)} + Y_{MF}^{(1)} > K^{(21)} \\
 (1,1) \rightarrow (0,1): & \quad C_{ON}^{(1)} + 4C_{ON}^{(1)}f_{nn}^{(1)} + 2f_{nn}^{(1)} + Y_{MF}^{(1)} > K^{(21)} \\
 (0,1) \rightarrow (0,0): & \quad 1 + 2C_{ON}^{(1)}f_{nn}^{(1)} + 4f_{nn}^{(1)} + Y_{MF}^{(1)} < K^{(21)} \\
 (0,0)_B \rightarrow (0,0): & \quad 1 + 6f_{nn}^{(1)} + Y_{MF}^{(1)} < K^{(21)}
 \end{aligned} \tag{S5.44}$$

Since  $f_{nn}^{(1)} < 1$  (i.e. the interaction of a cell with itself should be larger than that with its nearest neighbor), we can simplify the equations to account for redundancy. After some

algebraic manipulations, this reduces the conditions to the following set of inequalities:

$$\begin{aligned}
1 + 2C_{ON}^{(1)}f_{nn}^{(1)} + 4f_{nn}^{(1)} + Y_{MF}^{(1)} &< K^{(21)} \\
C_{ON}^{(1)} + 4C_{ON}^{(1)}f_{nn}^{(1)} + 2f_{nn}^{(1)} + Y_{MF}^{(1)} &> K^{(21)} \\
1 + 2C_{ON}^{(2)}f_{nn}^{(2)} + 4f_{nn}^{(2)} + Y_{MF}^{(2)} &< K^{(12)} \\
C_{ON}^{(2)} + 4C_{ON}^{(2)}f_{nn}^{(2)} + 2f_{nn}^{(2)} + Y_{MF}^{(2)} &> K^{(12)} \\
1 + 6f_{nn}^{(1)} + Y_{MF}^{(1)} &< K^{(11)} \\
1 + 2C_{ON}^{(1)}f_{nn}^{(1)} + 4f_{nn}^{(1)} + Y_{MF}^{(1)} &> K^{(11)}
\end{aligned} \tag{S5.45}$$

Note that for this particular example, the constraints reduce to a simple set of constraints for each of the three interactions in the system. Namely, the first two inequalities involve only parameters that affect the interaction  $2 \leftarrow 1$ , e.g.  $C_{ON}^{(1)}$  and  $K^{(21)}$ , whereas the second and third pair involve only the interactions  $1 \leftarrow 2$  and  $1 \leftarrow 1$  respectively. This does not need to be the case in general, since genes which are regulated by both genes will produce coupled constraints in terms of both interactions. Hence, in this case we can recast the conditions into a concise set of equations:

$$\begin{aligned}
K_{\min}^{(1,1)}(C_{ON}^{(1)}) &\leq K^{(1,1)} \leq K_{\max}^{(1,1)}(C_{ON}^{(1)}), \\
K_{\min}^{(1,2)}(C_{ON}^{(2)}) &\leq K^{(1,2)} \leq K_{\max}^{(1,2)}(C_{ON}^{(2)}), \\
K_{\min}^{(2,1)}(C_{ON}^{(1)}) &\leq K^{(2,1)} \leq K_{\max}^{(2,1)}(C_{ON}^{(1)})
\end{aligned} \tag{S5.46}$$

where the  $K_{\min}^{(i,j)}$  and  $K_{\max}^{(i,j)}$  are functions of  $C_{ON}^{(j)}$  defining the minimal and maximal possible values of  $K^{(i,j)}$ . Explicitly, we have

$$\begin{aligned}
K_{\min}^{(1,1)}(C_{ON}^{(1)}) &= 1 + 6f_{nn}^{(1)} + Y_{MF}^{(1)}, \\
K_{\max}^{(1,1)}(C_{ON}^{(1)}) &= 1 + 2C_{ON}^{(1)}f_{nn}^{(1)} + 4f_{nn}^{(1)} + Y_{MF}^{(1)}, \\
K_{\min}^{(1,2)}(C_{ON}^{(2)}) &= 1 + 2C_{ON}^{(2)}f_{nn}^{(2)} + 4f_{nn}^{(2)} + Y_{MF}^{(2)}, \\
K_{\max}^{(1,2)}(C_{ON}^{(2)}) &= C_{ON}^{(2)} + 4C_{ON}^{(2)}f_{nn}^{(2)} + 2f_{nn}^{(2)} + Y_{MF}^{(2)}, \\
K_{\min}^{(2,1)}(C_{ON}^{(1)}) &= 1 + 2C_{ON}^{(1)}f_{nn}^{(1)} + 4f_{nn}^{(1)} + Y_{MF}^{(1)}, \\
K_{\max}^{(2,1)}(C_{ON}^{(1)}) &= C_{ON}^{(1)} + 4C_{ON}^{(1)}f_{nn}^{(1)} + 2f_{nn}^{(1)} + Y_{MF}^{(1)}.
\end{aligned} \tag{S5.47}$$

Note also that  $Y_{MF}^{(i)}$  is a linear function of  $C_{ON}^{(i)}$  (Equation S5.39), so that these constraints reduce to linear relations between  $C_{ON}^{(i)}$  and  $K^{(i,j)}$  for each interaction  $i \leftarrow j$ . The values  $(C_{ON}^{(i)}, K^{(i,j)})$  together determine the relative strength of this interaction, and we find that its strength is constrained by two inequalities that determine a reduced but unbounded region of phase space. The boundaries of these regions together with the predicted TW

conditions are plotted in Figure S5.9B. This gives an alternative view of the parameter sets that can support TWs next to the spider charts, which only show that most parameters can span several orders of magnitude but do not directly reveal the structure of the set of TW parameters. In contrast, the analytic result reveals that each of the circuit parameters ( $C_{ON}^{(i)}, K^{(ij)}$ ) is unbounded from above, but is confined to a region such that each interaction can be neither too strong ( $C_{ON}^{(i)} \gg K^{(ij)}$ ) nor too weak ( $C_{ON}^{(i)} \ll K^{(ij)}$ ), except in the case of the self-activation loop where we tend to have  $C_{ON}^{(i)} \gg K^{(ij)}$ . Similar results to Figure S5.9B are obtained for the other networks that can support TWs. The set of inequalities (Equation S5.46-S5.47) also allows us to analytically calculate traveling wave robustness, as will be discussed in Section S5.5.4.

#### PERFORMANCE OF THE ANALYTIC FRAMEWORK

To assess the validity of the analytic framework derived in the previous sections, we directly compared the predictions from the theory to actual simulations of the waves. We quantified the degree to which these results match and considered the accuracy of the main approximation in the analytic framework, the nearest-neighbor approximation (Equations S5.36 and S5.39).

**Computational search for traveling waves** We verified with our analytic approach that the above wave forms are indeed the only possible wave forms for two-gene networks. To this end, we screened a large number of parameter sets for all distinct two-gene networks. Specifically, we checked the six conditions Equation S5.33 for wave propagation for a total of  $10^6$  parameter sets for each network. The parameter sets were generated by Latin hypercube sampling over all non-zero  $C_{ON}^{(i)}$  and  $K^{(ij)}$  parameters. We considered a network to be capable of generating a wave if for at least one of the  $10^6$  parameter sets all the conditions for traveling wave propagation were fulfilled. The results were consistent among the three types of waves (plane, with inward bends, with outward bends) that we examined: in all three cases exactly the same results were found.

**Statistical measures for performance** The performance of the analytic method we derived is determined by how well it predicts the conditions under which traveling waves can propagate. We can view our analytic theory as a binary classifier that predicts for a given gene network and given set of parameters whether TWs can propagate. The theory takes as input a set of parameters and gives as output a binary prediction about whether a TW can propagate or not. As such, we quantified its performance using well-established concepts for evaluating classifiers from machine learning. In particular, we look at the *accuracy*, *precision* and *recall* of the predictor for all the six cellular dialogues

and corresponding waveforms we found. These are defined as

$$\begin{aligned} \text{accuracy} &= \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \\ \text{precision} &= \frac{\text{TP}}{\text{TP} + \text{FP}} \\ \text{recall} &= \frac{\text{TP}}{\text{TP} + \text{FN}} \end{aligned}$$

Here TP = true positives, TN = true negatives, FP = false positives, FN = false negatives. True positives are parameter sets for which the TW propagates according to both theory and simulation. True negatives are parameter sets for which according to both theory and simulation TWs cannot propagate. False positives are predicted to be capable of sustaining TWs by the theory, but turn out not to do so in an actual simulation. False negatives are parameter sets that are capable of propagating TWs, but are missed by the theory.

5

**Assessment of the analytic framework (Figure S5.9)** For all of the networks that yielded waves, we find that the theory correctly predicts plane TWs to an extremely high degree of accuracy, close to 100% (Figure S5.9A). This means that the theory correctly predicts whether a wave can or cannot propagate in almost all cases. In contrast, the precision and recall take slightly lower scores, with a precision is between roughly 0.6-0.8 and a recall in the range of roughly 0.5-0.7. The interpretation is that roughly 60-80% of conditions predicted to allow TW propagation are indeed ones that can propagate a TW in an exact simulation, and that 50-70% of the all the conditions for TW propagation are correctly identified by the classifier. These lower values are caused by the low number of actual positives (conditions under which a TW can propagate), which is low compared to the total number of parameters we examined. This means that the few incorrect predictions that arise from the approximation have a relatively large impact on these performance metrics.

Overall, we thus obtain a good estimator for traveling wave propagation that is accurate except near the boundary of the regions permitting wave propagation. This also become apparent when we plot the interaction parameters of the predicted and actual waves together with the theoretical bounds (Figure S5.9B). This shows that the false predictions (false positives and false negatives) are mainly due to slight misestimations of the boundaries of the regions allowing for TW propagation. In particular, both the upper bound and lower bound for  $K^{(ij)}$  are slightly underestimated, meaning that the estimations for the mean-field contribution  $Y_{MF}^{(i)}$  are underestimated. This is evident from the fact that most false positives are near the lower bound for  $K^{(ij)}$  and most false negatives are near the upper bound for  $K^{(ij)}$ .

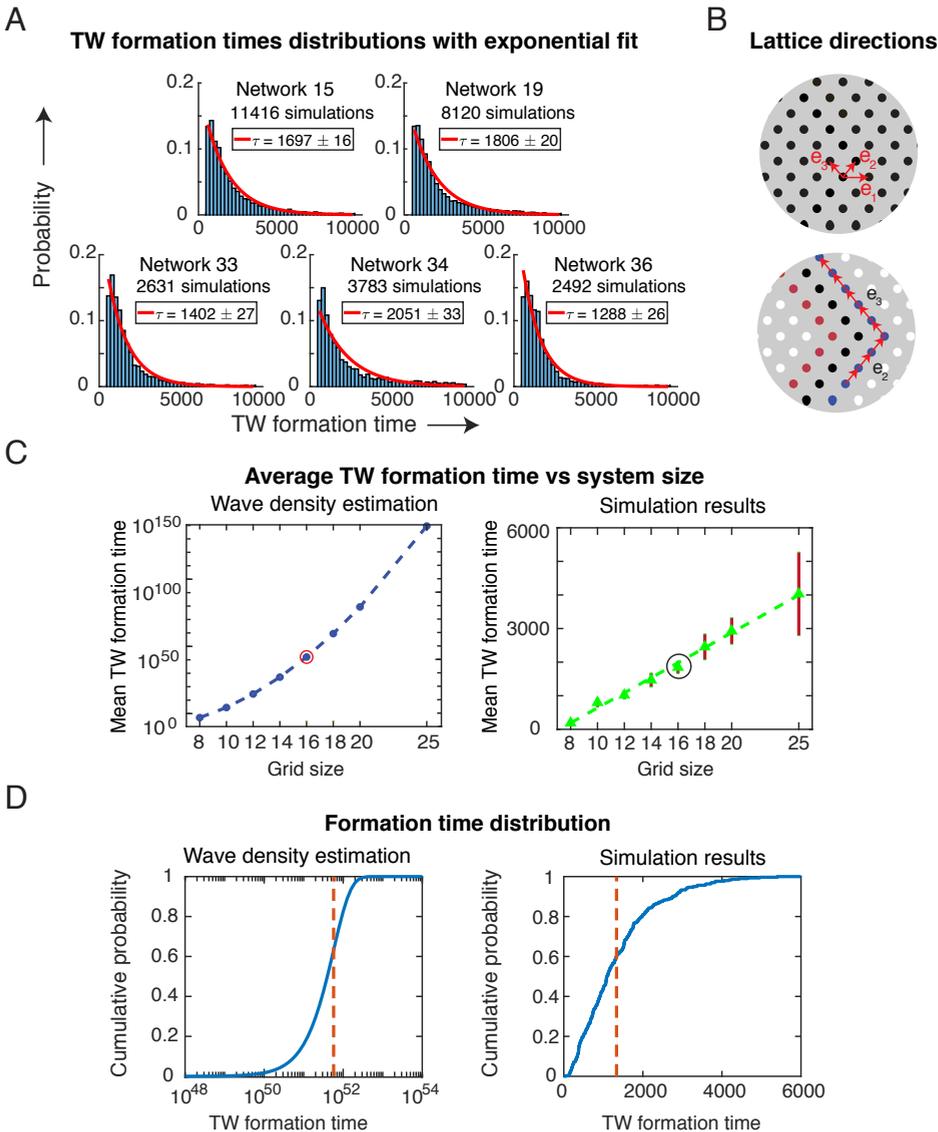
**Validity of the nearest neighbor approximation** The accuracy of the nearest neighbor approximation depends on how much of the total interaction the nearest neighbors capture. The more the nearest neighbors contribute to the total interaction strength, the more accurate the approximation is. This is because all deviations between the theory and the exact model come from the mean-field approximation, which has only a marginal contribution if the sensed concentration is mostly due to the cell itself and its nearest neighbors. We can quantify this by comparing  $f_{NN}^{(i)} \equiv 6f_{nn}^{(i)}$  (since there are six nearest neighbors) to the total interaction strength  $f_N^{(i)}$  for signaling molecule  $i$ . If  $f_{NN}^{(i)} \approx f_N$ , then the cells beyond the direct neighbors have only marginal influence on the concentration a cell senses. However, if  $f_{NN}^{(i)} \ll f_N$ , then the nearest neighbor approximation will be comparatively inaccurate, because we take into account the rest of the cells in an averaged manner only and neglect their spatial positions.

Note that  $f_{NN}^{(i)}/f_N$  depends on the parameters  $N$ ,  $a_0$  and  $\lambda^{(i)}$ . By examining how this quantity depends on these parameters, we get a picture of when the NNA is most accurate. For weak interaction (high  $a_0$ ), the nearest neighbor approximation matches closely with the actual system. For stronger interaction, the nearest neighbor approximation becomes worse (Figure S5.9C). In this case, one possible solution would be to extend the analysis to next-to-nearest neighbors, as we will discuss in the next part. In contrast, the ratio is hardly dependent on system size  $N$  and diffusion length  $\lambda^{(i)}$  (Figure S5.9C). Finally, longer diffusion length implies comparatively more influence from cells further away, leading to less accuracy for the nearest-neighbor approximation. However, this effect is also weak, accounting for less than 30% variation in interaction strength.

We also considered how taking next-to-nearest neighbors into account improved the accuracy of the analytic approach (Figure S5.9C – dotted lines). The contribution from next-to-nearest neighbors can be quantified by an interaction parameter  $f_{NNN}^{(i)} \equiv 4f^{(i)}(\sqrt{3}a_0) + 8f^{(i)}(2a_0)$ , which takes into account the interaction with the twelve cells in the second layer surrounding a cell on a hexagonal lattice. There are four cells at a distance of  $\sqrt{3}a_0$  and eight cells at a distance of  $2a_0$  in this layer. We find that total contribution to the interaction strength from the two layers of cells closest to a given cell,  $(f_{NN}^{(i)} + f_{NNN}^{(i)})$  indeed would improve accuracy, but the effects become less significant as  $a_0$  becomes larger. Thus, extending our framework to include next-to-nearest neighbors would improve the accuracy of our method, but as a NNA already perform well, it seems unnecessary to make our framework more complicated.

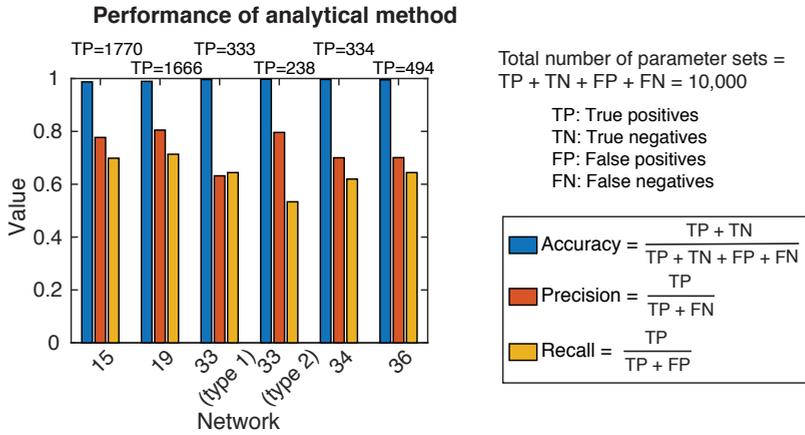
## OSCILLATORY TRAVELING WAVES

So far, we focused on traveling waves that are characterized by a propagating pattern on a fixed background. However, we also observed a variety of dynamic spatial patterns with oscillatory background cells in Networks 16, 20 and 43 (Figure S5.6A). In particular, oscillatory traveling waves (Section S5.5.1) form a subset that we can analyze using our framework. Here, we outline how to adapt our framework to the analysis of these dynamic patterns. From simulations, we observe that the oscillations have period 3 and always follows a fixed pattern. Using the definition of the Exterior (E), Front (F), Middle (M) and Back (B) states of a wave (S5.5.3 and Figure 5.4), we can trace out how these states transition on a state diagram (Section S5.5.2). We found that the cells of oscillatory waves follow a fixed pattern of cell state transitions (Figure S5.6D). Networks 16 and 20 each have one distinct state diagram, and network 43 can generate waves that follow either of the two state diagrams (Figure S5.6B). Each of the states undergoes a separate period 3 oscillation, but together they follow an regular pattern on the state diagram. At the transition with the dotted lines, the wave moves one step further (i.e., the entire pattern not only oscillates but also moves one cell layer ahead). This occurs for transitions where both genes are switched, either between (0,0) and (1,1) or between (0,1) and (1,0). Waves in network 43 can follow either the transitions of network 16 or those of network 20, or show more complicated patterns which fall outside these two standard cases. Hence, by imposing each of the transitions on either the same group of cells (when the wave oscillates but not moves) or a neighboring group of cells (when the wave translates), we could derive a more complicated set of constraints for the propagation of such waves if necessary.

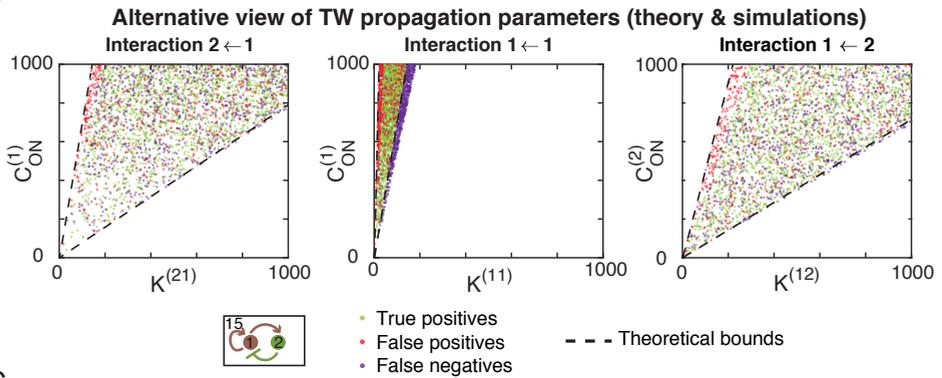


**Figure S5.8: Statistics on how long cells take to form traveling waves (TW).** (Related to Figure 5.5). (A) TW formation-time distributions from simulations (shown together in Figure 5.5G) are fitted by exponential functions, with  $\tau$  being the expectation value of the fitted exponential distribution. (B–D) Analytic calculation reveals that TW formation times do not follow an entirely random process, i.e. one where each next system state is randomly drawn from the set of all states, as one might suspect based on the chaotic appearance of the dynamics and the exponentially distributed formation times (see Section S5.5.3). (B) Constructions used in the calculation of the abundance of traveling waves in the system (see Section S5.5.3). (Left) Directions on the lattice. (Right) Sketch of the construction used to characterize a single wave. By counting all ways of traversing the lattice, subject to certain constraints, we obtain an estimate of the number of forms of traveling waves of a given type. (C) Average TW formation time estimated from the wave density calculation (left) at different grid sizes, compared with the empirical findings from exact simulations (right). Averages are taken over all self-organized TWs among 300 simulations per grid size, at fixed parameter values. Error bars represent s.e.m. The highlighted data point is the grid size used in (D). (D) The cumulative distribution of TW formation times according to the wave density estimation (top) and from simulations (bottom; from simulation set for grid size 16 also used in C). The red dotted line represents the average formation time.

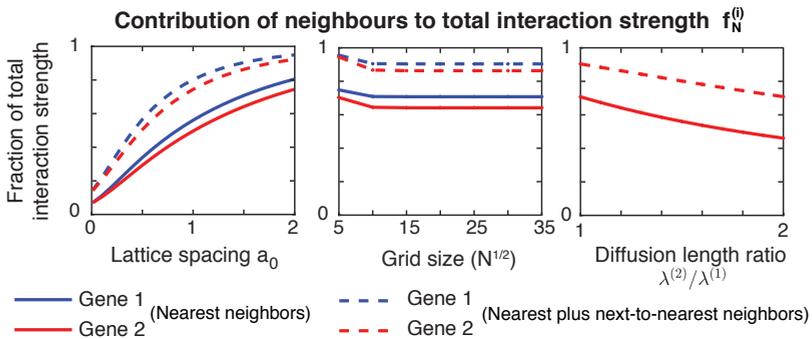
A



B



C



**Figure S5.9 (preceding page): Comparison between simulations and analytic theory for traveling wave propagation (also see Section S5.5.3. (Related to Figure 5.4)** (A) We used concepts from machine learning to quantify the performance of our analytic theory. Specifically, when viewed as a binary classifier, the analytic theory makes binary (yes or no) predictions about whether a parameter set is capable of propagating traveling waves. We compare these predictions with actual simulations to determine whether they are correct or false. The performance of the theory can then be quantified in terms of concepts such as accuracy, precision and recall (see Section S5.5.3). (B) Two-dimensional projections of the parameter sets that are capable of propagating traveling waves according to the analytic theory and exact simulations (see Figure 5.4F for alternate representations in terms of radar charts). Since there are six varying parameters for each parameter set, we projected the parameter sets onto two-dimensional spaces spanned by the two parameters describing the strength of each interaction – the threshold  $K^{(ij)}$  and the maximum secretion rate  $C_{ON}^{(j)}$ . We plot the data points classified as true positives, false positives and false negatives (see Section S5.5.3), but leave out the true negatives, which are the parameter sets which are correctly predicted to be incapable of sustaining traveling waves. (C) Contribution of nearest-neighbors ( $f_{nn}$ ) and next-to-nearest neighbors ( $f_{nnn}$ ) to the total interaction strength (see Section S5.5.3), as a function of the lattice spacing, the grid size and the ratio between the diffusion lengths. We plot the contribution from nearest neighbors and next-to-nearest neighbors as a fraction of the total interaction strength.

#### S5.5.4. ROBUSTNESS AND RELIABILITY OF TRAVELING WAVES

##### ROBUSTNESS (FIGURE S5.10A–C)

Biological robustness is typically referred to as the ability of a biological system to adapt to environmental perturbations by maintaining its function [Kitano, 2004]. Control mechanisms such as feedback loops may play a role in maintaining robustness. In our system, different time scales allow us to study the concept of robustness at different levels. The dynamics of the parameters of the system occurs at an evolutionary time scale (unless the experimentalist intervenes), while the dynamics of the gene expression happens on a much shorter time scale (minutes to hours) and the dynamics of the signaling factors occurs on an even faster time scale. As such, we may consider perturbations at each of these levels of description to see how they affect the system's ability to perform a certain function — which in this case means it's ability to generate patterns. In this paper, we considered the system's response to changes in parameter values. Since we assume the parameters to stay constant during the entire simulation, we will use robustness as a static quantity obtained by comparing simulations at different sets of fixed parameters. Specifically, we considered the robustness of traveling waves for two different situations. We considered the robustness of TW formation — how changing parameters impacts the system's ability to self-organize into a TW, as well as the robustness of TW propagation — how parameters influence the ability of an already formed TW to continue propagating. We quantified the robustness in both cases by the fraction of parameter sets, or Q-value, that can generate or propagate a TW [Von Dassow et al., 2000; Ma et al., 2006]. In the absence of further information about the parameters, this tells us how likely it is to find parameters which are compatible with a certain property or behavior of the system — formation or propagation of TWs in our case.

**Normalized Q-value** The Q-values we obtain as a fraction of parameter set compatible with TWs depend on the number of parameters  $m$  we sample over for each network. These values will tend to be higher for networks with fewer parameters than for networks with higher number of parameters. One method used to correct for this is to take the  $m$ -th root of the Q-value [Von Dassow et al., 2000]. We will call this the “normalized Q-value”. This value represents the chance for each of the  $m$  parameters to be compatible with TW formation over a specified range of values.

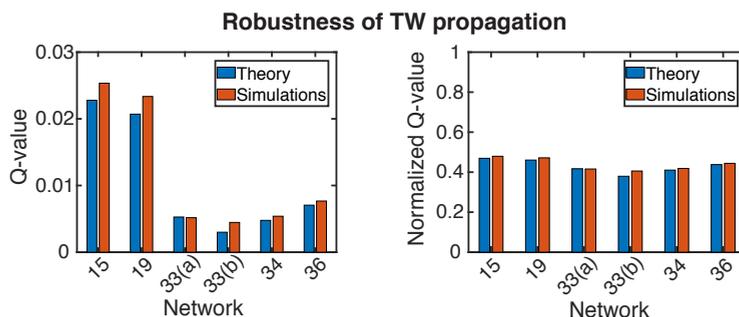
**Calculation of Q-values from simulation data** In the absence of any predictive theory, the phase space volume compatible with TWs can only be estimated through drawing random samples from the parameter space and determining whether TWs can form or propagate for each of these samples. Note that in principle our parameters are unbound, i.e.  $K^{(ij)} \geq 1$  and  $C_{ON}^{(j)} \geq 1$  with no upper bound. To calculate the robustness, we there-

fore specified a finite region defined by  $1 \leq K^{(ij)} \leq L$  and  $1 \leq C_{ON}^{(j)} \leq L$  for each signaling molecule  $j$  and each interaction  $i \leftarrow j$  (neglect the parameters for non-existent interactions). In practice, we took  $L = 1000$  everywhere. We then used Latin hypercube sampling to generate a large number of parameter sets and tested whether TWs could form or propagate for each of the parameter sets.

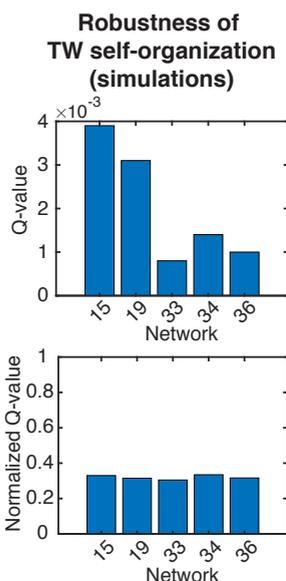
For TW formation, we tested how likely it is to find self-organization of TWs in the five networks for which we found self-organized TWs (Figure 5.3D). We used the same 10,000 sampled parameter sets as were used to generate the initial network classification (Figures 5.3 and S5.1). For each parameter set, we considered it to be capable of self-organizing TWs if at least one simulation (out of 10 runs per parameter set) led to a self-organized TW. The Q-value we obtained for TW formation in this way is of the order of  $10^{-3}$  (Figure S5.10B — upper figure). Alternatively, after correcting for the number of parameters, the normalized Q-value corresponds to a randomly generated parameter having around 30% chance of taking a value compatible with TW formation across a 1,000-fold range for each parameter (Figure S5.10B — lower figure). The Q-values obtained in this way are in fact lower estimates as we perform only a finite number of simulations and would be higher if we could screen over all possible initial states (in which case the Q-values for TW formation and TW propagation would coincide, since we would also include the final pattern as initial state). Nevertheless, this approach mirrors the situation in wet lab experiments, where can only test a finite number of replicates before concluding that a particular result is highly unlikely to be reached.

For TW propagation, we tested the two types of TWs we found (Figure 5.4D) for each of the networks in which we found them. We used the same data as obtained from Latin hypercube sampling which we used to quantify the performance of our analytic predictor (Figure S5.9), as this contains precise information on whether each parameter set should be able to propagate TWs according to both the theoretical prediction as well as explicit simulations. This gave higher Q-values, in the order of  $10^{-2}$  for TW propagation (Fig. S5.10A — left figure), corresponding to normalized Q-values of about 40 – 50% for each parameter to be compatible (Figure S5.10A — right figure). In comparison, the robustness of the *Drosophila* segment polarity gene network was quantified for a network with a far larger set of parameters, for which the authors found Q-values corresponding to normalized values of about 80 – 90% for each parameter [Von Dassow, 2000].

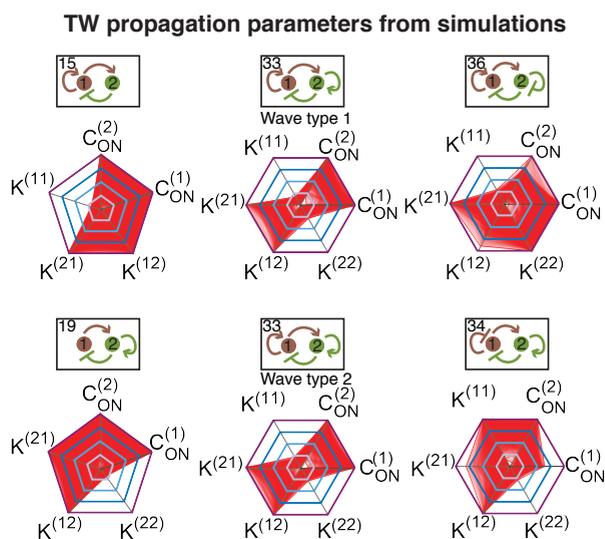
A



B



C



**Figure S5.10: Robustness of traveling waves (TWs), measured by the fraction of parameter sets capable of propagating traveling waves (also see Section S5.5.4). (Related to Figures 5.4 and 5.5)** (A) We defined robustness as the fraction of parameter sets that were capable of propagating a TW (“Q-value” – see Section S5.5.4). Here we show the robustness of TW propagation for each cellular dialogue that is capable of generating TWs. Networks 33(a) and 33(b) refer to the two types of TWs that cellular dialogue 33 can generate (see Figure 5.4D). The normalized Q-value considers the number of parameters for each parameter set and can be interpreted as the probability that a single random draw of each parameter value yields a TW (see Section S5.5.4). Results are based on testing  $10^6$  randomly generated parameter sets obtained from Latin hypercube sampling for both theory and simulations (see Section S5.4.2). (B) Robustness of TW self-organization from random initial states. Results are based on testing  $10^4$  parameter sets, with 10 simulations for each parameter set. (C) Radar charts or spider charts for the parameter sets for which TWs propagate as found in simulations (compare with theoretical results in Figure 5.4F).

#### ROBUSTNESS: ANALYTIC CALCULATION (FIGURE S5.11)

We can also obtain an estimate for the robustness of TWs through the analytically derived conditions (Section S5.5.3), which we will apply to the explicitly derived conditions

for network 15 (Equation S5.46). The derived inequalities define a region  $\mathcal{U} \subset \mathcal{P}_L$  that is compatible with TW propagation. The volume of this region,  $V(\mathcal{U})$ , in relation to the total phase space volume  $V(\mathcal{P}_L)$  defines the robustness, i.e. we can express the Q-value as

$$Q = \frac{V(\mathcal{U})}{V(\mathcal{P}_L)}. \quad (\text{S5.48})$$

The volume over the region compatible with traveling waves can be expressed as an integral over  $\mathcal{P}_L$ :

$$V(\mathcal{U}) = \int_1^L dC_{ON}^{(1)} \int_1^L dC_{ON}^{(2)} \int_1^L dK^{(1,1)} \int_1^L dK^{(1,2)} \int_1^L dK^{(2,1)} \mathbb{1}_{TW}, \quad (\text{S5.49})$$

where the function  $\mathbb{1}_{TW}$  takes values 1 on the domain for which the inequalities are satisfied and 0 elsewhere. We have shown that the derived conditions can be reduced to sets of independent conditions on  $K^{(ij)}$  and  $C_{ON}^{(j)}$  for each interaction  $i \leftarrow j$ . This means that the integral decomposes into two separate integrals

$$\begin{aligned} V(\mathcal{U}) &= V^{(1)} V^{(2)} \\ V^{(1)} &= \int_1^L dC_{ON}^{(1)} \int_1^L dK^{(1,1)} \int_1^L dK^{(2,1)} \mathbb{1}_{TW} \\ V^{(2)} &= \int_1^L dC_{ON}^{(2)} \int_1^L dK^{(1,2)} \mathbb{1}_{TW} \end{aligned} \quad (\text{S5.50})$$

To evaluate  $V^{(1)}$  and  $V^{(2)}$ , one must take into account the various ways in which the borders defined by the inequalities intersect with the boundaries of the box. The integrals over  $C_{ON}^{(j)}$  can then be split into integral over the various regions defined by these intersections. Concretely, let us define  $C_{min}^{(ij)}$  and  $C_{max}^{(ij)}$  as values of  $C_{ON}^{(j)}$  at which  $K_{min}^{(ij)} = L$  and  $K_{max}^{(ij)} = L$  respectively. Note that  $C_{min}^{(ij)} > C_{max}^{(ij)}$  by definition. For  $V^{(2)}$ , we have three different cases to distinguish: If  $L < C_{max}^{(12)} < C_{min}^{(12)}$ , we have a single integral

$$V^{(2)} = \int_1^L dC_{ON}^{(2)} (K_{max}^{(12)} - K_{min}^{(12)})$$

If  $C_{max}^{(12)} < L < C_{min}^{(12)}$ , we have to split up the integral into two parts:

$$V^{(2)} = \int_1^{C_{min}^{(12)}} dC_{ON}^{(2)} (K_{max}^{(12)} - K_{min}^{(12)}) + \int_{C_{min}^{(12)}}^L dC_{ON}^{(2)} (L - K_{min}^{(12)})$$

Finally, if  $C_{max}^{(12)} < C_{min}^{(12)} < L$ , we have to split up the integral into three parts, where the integrand over the third part is zero:

$$V^{(2)} = \int_1^{C_{min}^{(12)}} dC_{ON}^{(2)} \left( K_{max}^{(12)} - K_{min}^{(12)} \right) + \int_{C_{min}^{(12)}}^{C_{max}^{(12)}} dC_{ON}^{(1)} \left( L - K_{min}^{(12)} \right) + \int_{C_{max}^{(12)}}^L dC_{ON}^{(1)} 0$$

In the case of  $V^{(1)}$ , the procedure is a bit more involved, but can be summarized as follows: we first order the values  $C_{min}^{(11)}, C_{max}^{(11)}, C_{min}^{(21)}, C_{max}^{(21)}$ , so that we can split the integration domain of  $C_{ON}^{(j)}$  up into  $m \leq 4$  parts  $[b_0 = 1, b_1], [b_1, b_2], \dots, [b_{m-1}, b_m = L]$  by taking all of the ordered values which are less than or equal to  $L$ . For each integration segment  $[b_{s-1}, b_s]$ , we then distinguish between the same three cases as in the previous calculation, but now with an integrand that is the product of two factors, which we will denote  $F_s^{(11)}$  and  $F_s^{(21)}$  for now (for  $K^{(11)}$  and  $K^{(12)}$  and segment number  $s$ ). The cases to distinguish are then similar:

1. If  $b_i < C_{max}^{(i1)} < C_{min}^{(i1)}$ , then  $F_s^{(i1)} = \left( K_{max}^{(1j)} - K_{min}^{(1j)} \right)$ .
2. If  $C_{max}^{(i1)} < b_i < C_{min}^{(i1)}$ , then  $F_s^{(i1)} = \left( L - K_{min}^{(1j)} \right)$ .
3. If  $C_{max}^{(i1)} < C_{min}^{(i1)} < b_i$ , then  $F_s^{(i1)} = 0$ .

The integral is then calculated as

$$V^{(2)} = \sum_{i=0}^m \int_{b_i}^{b_{i+1}} dC_{ON}^{(1)} F_s^{(11)} F_s^{(21)}. \quad (S5.51)$$

**Maximal robustness** Equipped with an explicit expression for the robustness of TWs, we can now study how the robustness changes with parameters, and in particular how to maximize the robustness by tuning parameters of the system. We quantified the robustness through a Q-value calculated over parameters describing the parameters of the gene circuit ( $K^{(ij)}$  and  $C_{ON}^{(j)}$ ), but this value will depend on other parameters of the system. In particular, the expressions of the boundaries  $C_{min}^{(ij)}$  and  $C_{max}^{(ij)}$  already reveal that the Q-value directly depends on  $f_{nn}^{(i)}$  and  $f_N^{(i)}$  (Equations S5.47 and S5.39). These quantities in turn depend on  $N$ ,  $a_0$  and  $\lambda^{(i)}$  ( $i = 1, 2$ ). Hence, we will study how the Q-value changes with these four parameters, and in particular whether there are global maxima for the Q-value as a function of these parameters. Note that we can reduce the number of parameters to three by normalizing the diffusion lengths,  $\lambda^{(i)}$  (see STAR Methods).

We first numerically sampled over  $N, a_0$  and found a consistent dependence of the robustness on  $a_0$  for different values of the system size (Figure S5.11A-B). The Q-value first rapidly increases with  $a_0$ , reaches a maximum and then decays towards zero (Figure S5.11B). Intuitively, this trend can be understood as a balance between only self-signaling (at large  $a_0$ ) and excessive communication (at small  $a_0$ ). Namely, at large  $a_0$ , the cells only sense their own signaling molecules. But we have seen that traveling waves are emergent phenomena that rely on cell signaling to realize different dynamics for cells with the same states, based on their location in relation to the wave. Thus, as  $a_0 \rightarrow \infty$ , necessarily the robustness of the waves goes to zero. Mathematically, this is evident from the fact that  $f_{nn} \rightarrow 0$  as  $a_0 \rightarrow \infty$ , so that locally different neighborhoods now exhibit the same dynamics. Conversely, when  $a_0$  approaches zero, the Q-value also goes to zero (for large system sizes at least). This is likely because the interaction between the cells becomes too strong, impeding transitions required to turn off genes, which can only occur if the sensed concentrations are low enough.

For very small system sizes (grid size  $\lesssim 10$ ), the trend differs at low values of  $a_0$ , where interactions beyond nearest-neighbors become important (Figure S5.11D). But since traveling waves are only of interest in large enough systems, we will not consider their behavior in very small systems. Conversely, in the  $N \rightarrow \infty$  limit,  $f_N$  approaches a constant value and therefore the Q-values also converge to a constant, as is already evident from the trend at the largest grid sizes shown in Figure S5.11D.

We then examined the effect of varying the diffusion lengths. We will work with the normalized lengths  $l^{(1)}, l^{(2)}$ , so that our results are independent of  $a_0$ . Since we have previously seen that the results have only weak dependence on  $N$ , effectively we can consider the robustness as a function of these two parameters only. A numerical screening across these two parameters reveals the presence of a single global maximum for the Q-value (Figure S5.11A), at a value of  $l^{(1)} \approx 2.2, l^{(2)} \approx 0.7$ . This suggests that the robustness of traveling waves can be optimized by appropriately choosing the signaling molecules (in particular their diffusion lengths) the cells use to communicate with. In the following, we will try to explain the presence of this single maximum through a more detailed analysis.

**Area fractions** Whereas the Q-value is a measure for overall robustness, we can focus on individual interactions and look at how likely it is to find parameters compatible with TWs for each interaction. We first project the parameter set onto a 2D-plane described by  $K^{(ij)}, C_{ON}^{(j)}$  (for an interaction  $i \leftarrow j$ ), as these two parameters together specify the relative strength of the interaction. We then determine area spanned by  $K_{min}^{(ij)}$  and  $K_{max}^{(ij)}$  in this plane and calculate the fraction of this area with respect to the total phase space

area considered  $(L-1)^2$  (recall that we assumed that  $K^{(ij)}$  and  $C_{ON}^{(j)}$  span a similar range of values as they have the same units). Formally, we can therefore write

$$a^{(ij)}(L) = \frac{1}{(L-1)^2} \int_1^L dK^{(ij)} \int_1^L dC_{ON}^{(j)} \mathbb{1}_{TW}. \quad (\text{S5.52})$$

These area fractions represent the probability of randomly picking the right parameters for each interaction. We then calculated how these area fractions depend on the signaling lengths  $l^{(1)}, l^{(2)}$ . By plotting how these individual area fractions depend on these variables, we obtain a better understanding of how the maximum Q-value arises (Figure S5.11E). First, we note that since each area fraction depends only on one signaling length, we can study how they vary with this single parameter (Figure S5.11F). We see that the area fraction for  $i = 1, j = 1$  increases with  $l^{(1)}$ , whereas the area fractions for  $i = 1, j = 2$  and  $i = 2, j = 1$  first slightly increase and then decrease with  $l^{(2)}$  and  $l^{(1)}$  respectively. Although the Q-value does not directly decompose into a product of area fractions, we still expect it to be high if and only if all area fractions are relatively high. Thus, to obtain optimal robustness, for the second signaling molecule we simply have to tune its signaling length  $l^{(2)}$  to the maximum of the area fraction  $a^{(12)}$ . For the first signaling molecule, there is a competition between maximizing the area fractions of  $a^{(11)}$  and  $a^{(21)}$ , which generally have opposing trends. Hence, optimal robustness is likely to be found at an intermediate value of  $l^{(1)}$  that is neither too low or too high.

The trend in the area fractions can be further interpreted by explicitly examining the projected areas and analyzing limiting cases explicitly. We first note that in the limits  $l^{(1)} \rightarrow 0$  and  $l^{(1)} \rightarrow \infty$ , the interaction strength becomes

$$\begin{aligned} \lim_{l^{(i)} \rightarrow 0} f^{(i)}(\rho) &\rightarrow 0, \\ \lim_{l^{(i)} \rightarrow \infty} f^{(i)}(\rho) &\rightarrow \frac{r_{cell}}{\rho}, \end{aligned} \quad (\text{S5.53})$$

where  $\rho = r/a_0$  as we recall. Indeed,  $l^{(1)} \rightarrow 0$  means that the signaling molecules hardly diffuse anymore, so the interaction between cells becomes negligible. Conversely, if  $l^{(1)} \rightarrow \infty$ , it means that the signaling molecule is basically never broken down as it diffuses away from its source. In this case,  $f^{(i)}(\rho)$  reaches a constant that is still distance-dependent, since the same concentration is spread across a larger and larger area (or volume in 3D) as the molecules diffuse away from the source. The  $1/\rho$  dependence implies that the concentration on annulus of inner radius  $\rho$  and width  $d\rho$ ,  $f^{(i)}(\rho) \rho d\rho$ , is independent of distance.

With this knowledge, we are now in a position to understand the dependence of the area

fractions on the signaling lengths of the two molecules. Let us do this by considering each of the interactions separately.

First, consider the interaction  $1 \leftarrow 1$ . The bound  $K_{min}^{(11)}$  represents the lowest  $K^{(11)}$  value at which the background cells remain off, so the transition  $(0,0)_E \rightarrow (0,0)_E$  can occur. The bound  $K_{max}^{(11)}$  represents the highest value at which gene 1 can turn ON or remain ON, which is required for the transitions  $(0,0)_F \rightarrow (1,0)$  and  $(1,0) \rightarrow (1,1)$ .

1. In the limit  $l^{(1)} \rightarrow 0$ , we have  $K_{min}^{(11)} \rightarrow 1$  and  $K_{max}^{(11)} \rightarrow 1$ , and therefore the area fraction  $a^{(11)}$  goes to  $a^{(11)} \approx 0$  (left plot of Figure S5.11F). The system cannot simultaneously keep gene 1 off in the background state  $(0,0)$  and turn on gene 1 in the  $E_F$  cells in front of the wave.
2. In the limit  $l^{(1)} \rightarrow \infty$ , we have  $K_{min}^{(11)} \rightarrow 7 + Y_{MF}^{(1)}$  and  $K_{max}^{(11)} \rightarrow 5 + 2C_{ON}^{(1)} Y_{MF}^{(1)}$ , and therefore the area fraction  $a^{(11)}$  reaches a constant value (left plot of Figure S5.11F). Since both boundaries have relatively large slope in  $C_{ON}^{(1)}$ , the area is moderately small.

Then, consider the interaction  $1 \leftarrow 2$ . Since this is a repressive interaction, the role of the bounds have reversed.  $K_{min}^{(12)}$  represents the lowest value of  $K^{(12)}$  for which gene 1 can be ON (unrepressed), whereas  $K_{max}^{(12)}$  represents the highest value at which gene 1 can still be repressed.

1. In the limit  $l^{(2)} \rightarrow 0$ , we have  $K_{min}^{(12)} \rightarrow 1$  and  $K_{max}^{(12)} \rightarrow C_{ON}^{(2)}$ , and therefore the area fraction goes to  $a^{(11)} \approx 1/2$  (middle plot of Figure S5.11F). To turn or keep on gene 1,  $K^{(12)}$  cannot be too low. But this is required only when gene 2 is off (for the transitions  $(0,0)_F \rightarrow (1,0)$  and  $(1,0) \rightarrow (1,1)$ ), in this limit the sensed concentration of gene 2 is very low for these states. Conversely, to turn off gene 1 through repression by gene 2,  $K^{(12)}$  cannot be too high. However, since turning off is only required in states where gene 2 is ON (namely, for the transitions  $(1,1) \rightarrow (0,1)$  and  $(0,1) \rightarrow (0,0)$ ), and the cell senses a concentration  $C_{ON}^{(2)}$  by default in these states, the threshold  $K^{(12)} \geq C_{ON}^{(2)}$ . Together, this leaves a relatively large area permitted.
2. In the limit  $l^{(2)} \rightarrow \infty$ , we have  $K_{min}^{(12)} \rightarrow 5 + 2C_{ON}^{(2)} + Y_{MF}^{(2)}$  and  $K_{max}^{(12)} \rightarrow 2 + 5C_{ON}^{(2)} + Y_{MF}^{(2)}$ , and therefore the area fraction  $a^{(12)}$  reaches a constant value (middle plot of Figure S5.11F). Since both boundaries have relatively large slope in  $C_{ON}^{(2)}$ , the area is moderately small.

Altogether, this implies that the interaction  $1 \leftarrow 2$  should favor a relatively small diffusion length  $l^{(2)}$ , as is observed. The analysis does not explain the existence of a maximum between these extremes.

Finally, for the interaction  $2 \leftarrow 1$ , the equations are identical to those for  $1 \leftarrow 2$  and there-

fore the analysis is similar, with the difference that this is an activating interaction and therefore the interpretations of turning ON and OFF should be reversed.

In summary, although this technical analysis is rather involved, in essence the existence of a single maximum in robustness can be understood through a competition between self-interaction and neighbor interaction. Transitions which occur mostly due to self-interaction favor weak interaction with neighbors (low  $l^{(1)}$  and  $l^{(2)}$ ). Conversely, transitions that rely on neighbor interactions (e.g. neighbor-induced gene activation for the interaction) favor strong interaction or high signaling lengths. The competition between these two effects is responsible for creating a single optimal robustness in terms of the signaling lengths.

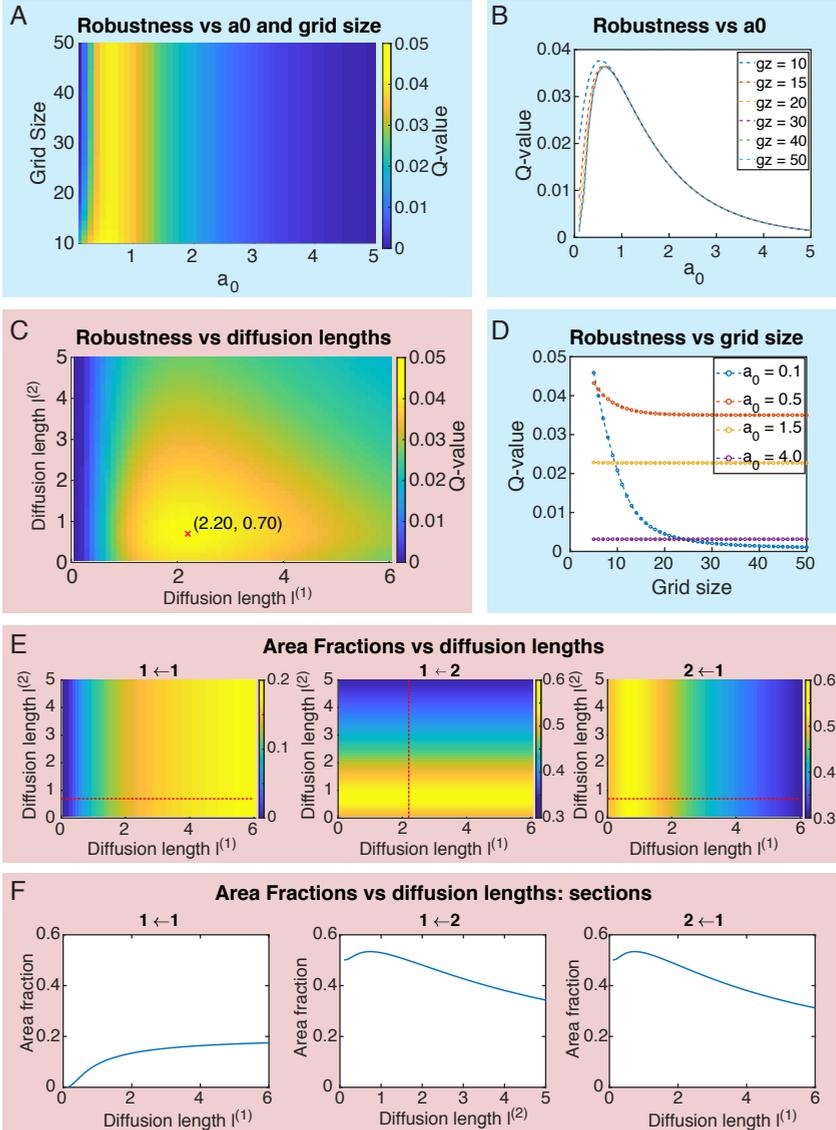
#### RELIABILITY (FIGURES 5.5F AND S5.12)

## 5

While robustness deals with the probability of finding parameter sets compatible with TWs, we can also ask what the chance of finding a TW is once the parameter set has been fixed. We define the *reliability* of TW formation as the percentage of simulations with varying initial conditions that generate TWs given a set fixed parameters. For each of the sets of parameters that yielded self-organized TWs, we determined the reliability by running a large set of simulations and counting in how many of those TWs spontaneously formed. Overall, we found an average reliability of 0.2-0.4 across all networks, indicating that we expect TWs to form in roughly 20-40% of the time for these parameter sets (Figure 5.5F). However, upon closer inspection we find that this average results from a considerable variability between different parameter sets, indicating that the precise choice of parameters has a large influence on the reliability of TW formation (Figure S5.12A). While for many parameter sets the reliability is exceedingly low (5-10%), there is a continuum of reliability values all the way up to about 80% (Figure S5.12A).

This finding raises the question of whether we could identify any source of this variability in reliability values between different parameter sets. To address this question, we took a large set of parameter sets ( $n = 2534$ ) capable of sustaining a TW once it has formed, as tested explicitly in simulations starting with a TW. For each of these parameter set, we then ran a large number of simulations (100) to see whether it could also self-organize into traveling waves if we set up the initial configuration to be random. Surprisingly, we found that a large set of these parameter sets did not yield self-organized waves at all (Figure S5.12B). This indicates the system may be able to propagate a pattern, but have only few ways of generating such a pattern in the first place. Between the parameter sets that were found to self-organized TWs, the reliability varies dramatically along a con-

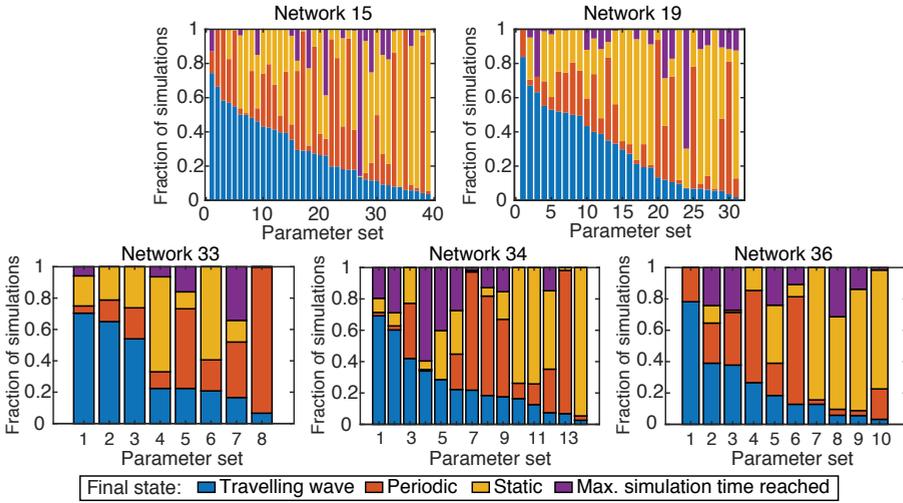
tinuum between 0 — virtually no simulations become TWs — to close to 1 — almost all simulations become TWs. When we plot the distribution of the 578 parameter sets found to generate TWs, we find that the probability to find a given reliability decays nearly monotonically (Figure S5.12C), indicating that parameter sets with higher reliability increasingly rare. However, when we examined the reliability of these different parameter sets as a function of the parameters, we observed no clear trend or correlation in two different projections of the parameter sets (Figure S5.12D-E). The parameters sets with extremely high reliability values are scattered around the entire region in which TWs are possible (Figure S5.12D). Furthermore, for any of the parameters, there are parameter sets with high reliability for both very high and very low values of that parameters, and the same applies to low reliability (Figure S5.12E).



**Figure S5.11: Analytic framework predicts robustness of wave formations (also see Section S5.5.4). (Related to Figure 5.5).** We used our analytic framework to compute how the robustness –defined as the Q-value (i.e., fraction of parameter sets that enable wave propagation)– varies with the number of cells in a population (light blue plots), nearest-neighbor distance  $a_0$  (i.e., lattice spacing), and diffusion lengths  $l^{(1)}, l^{(2)}$  (see S5.4.1) (pink plots). **(A–D)** Q-value as a function of these parameters. **(B)** and **(D)** are one-dimensional sections of the plot in **(A)**, obtained by fixing one of the two parameters while varying the other. **(E–F)** Area fractions are fractions of the two-dimensional parameter space spanned by  $K^{(i,j)}, C_{ON}^{(j)}$  for a single interaction  $i \leftarrow j$  that allow for wave propagation (also see Section S5.5.4). We show here how the area fractions vary with the diffusion lengths  $l^{(1)}, l^{(2)}$ . Red dotted lines in **(E)** show how we fixed one of these two parameters to obtain the plots in **(F)**.

A

**Reliability of TW formation - by network and parameter set**

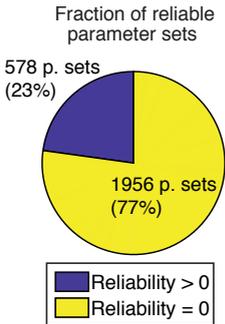


5

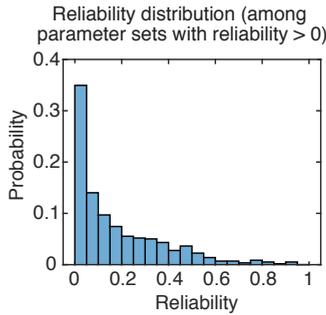
**Reliability of self-organization - parameter sets that can propagate TWs**

(Network 15, n=2534)

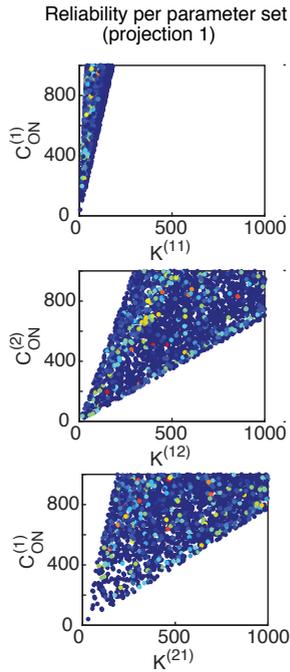
B



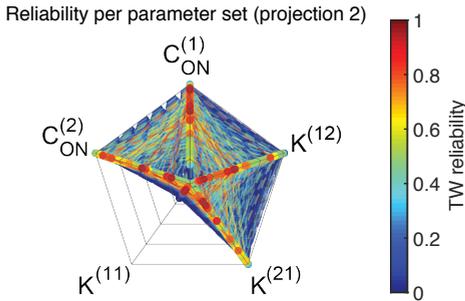
C



D



E



**Figure S5.12 (preceding page): Reliability of traveling wave (TW) formation (also see Section S5.5.4) (Related to Figure 5.5).** We defined “reliability of TW formation” as the fraction of simulations that start with disordered spatial-configurations which then form TWs (see Section S5.5.4). **(A)** Reliability values for all parameter sets that led to TWs, for each of the five cellular dialogues that can generate dynamic spatial-patterns. The parameter sets are ordered from those that generated the highest number of TWs to those that generated the lowest number of TWs (blue bars). We classified the other simulations as “other periodic patterns” (red), “static patterns” (yellow) and simulations that did not reach either a periodic or a static steady-state (purple). We performed 500 simulations with random initial spatial-configurations for each parameter set. (*Caption continued on next page.*) **(B-E)** Reliability of TW formation for a large set of parameters ( $n = 2534$ ) capable of propagating TWs (i.e., for which a TW initial state continued to propagate indefinitely). For each parameter set, we performed 100 simulations to test whether random initial conditions led to self-organization of TWs. **(B)** Fraction of parameter sets that yielded at least one self-organized TW. **(C)** Distribution of reliability values among the set of parameters with positive values for reliability. **(D-E)** Reliability shows no clear dependence on any of the five parameters that we varied. **(D)** Projection of the five-dimensional parameter sets on the two-dimensional parameter space, spanned by the two parameters that specify the strength of each interaction (see Figure S5.9B for details). Each dot represents one parameter set and the color represents the reliability (color bar shared with Figure S5.9B). **(E)** Spider chart projection. Each connected thread represents one parameter set, with the color of the thread representing the reliability for that parameter set.

## PERSISTENCE OF TRAVELING WAVES WITH NOISE (FIGURE S5.13)

Recall that we observe that traveling waves persist with complex elements up to a certain degree (Figure 5.6), but that decreasing Hill coefficient or increasing noise, lattice disorder or cell motility beyond certain thresholds will cause the waves to stop propagating. We obtained these results from running simulations with different values for the new parameters the extended model introduces (i.e. noise, Hill coefficient, etc.). While this is a feasible way to obtain information about persistence of TWs, ideally we would like to be able to estimate these effects without running any simulations. To this end, we derived an analytic method to compute the effect of noise on traveling wave persistence (Figure S5.13).

5

We study the propagation conditions derived in Section S5.5.3 under the influence of noise in the form of fluctuating thresholds  $K^{(ij)} \rightarrow K^{(ij)} + \delta K^{(ij)}$  with  $\delta K^{(ij)} \sim \mathcal{N}(0, \sigma K^{(ij)})$ , as defined previously. Under our wave decomposition scheme, we have six different transitions that need to be satisfied by a given number of cells. Consider the transition  $\alpha \rightarrow \beta = X(\alpha)$  for a transition of a wave state  $\alpha \in \{E_F, F, M, B, E_B, E\}$  to a cell state  $\beta = (\beta^{(1)}, \beta^{(2)}), \beta^{(i)} \in \{0, 1\}$ . The condition under which this transition occurs can be written as

$$\beta^{(i)} = \prod_j g^{(ij)}(X(\alpha)), \quad (\text{S5.54})$$

$$g_\alpha^{(ij)} = \begin{cases} \theta(Y_\alpha^{(j)} - K^{(ij)}) & M^{(ij)} = 1 \\ \theta(K^{(ij)} - Y_\alpha^{(j)}) & M^{(ij)} = -1 \\ 1 & M^{(ij)} = 0 \end{cases}. \quad (\text{S5.55})$$

Here  $g_\alpha^{(ij)} = g^{(ij)}(X(\alpha))$  and  $Y_\alpha = (Y_\alpha^{(1)}, Y_\alpha^{(2)})$  is the sensed concentration of the cell with wave state  $\alpha$ , which we calculate in the nearest-neighbor approximation (Section S5.5.3). The persistence of the TW requires that the transition conditions are met for each of the  $N$  cells of the system. We next derive the probability that this occurs for a given value of the noise strength  $\sigma$ .

First, we derive the probability that  $g^{(ij)}(X(\alpha)) = 1$  for any interaction and  $\sigma$ . The probability that this holds depends on the the interaction type specified by  $M^{(ij)}$ . For  $M^{(ij)} = 0$ , this condition is trivially met, so we consider  $M^{(ij)} \neq 0$ . We then have the general expres-

sion

$$\begin{aligned}
P(\mathbf{g}_\alpha^{(ij)} = 1) &= P\left(Y_\alpha^{(j)} - K^{(ij)} - \delta K^{(ij)} M^{(ij)} > 0\right) \\
&= \begin{cases} P\left(\delta K^{(ij)} < Y_\alpha^{(j)} - K^{(ij)}\right) & M^{(ij)} = 1 \\ P\left(\delta K^{(ij)} > Y_\alpha^{(j)} - K^{(ij)}\right) & M^{(ij)} = -1 \end{cases} \\
&= \begin{cases} D(Y_\alpha^{(j)} - K^{(ij)}; 0, \sigma K^{(ij)}) & M^{(ij)} = 1 \\ 1 - D(Y_\alpha^{(j)} - K^{(ij)}; 0, \sigma K^{(ij)}) & M^{(ij)} = -1 \end{cases} \\
&= \left(\frac{1 + M^{(ij)}}{2}\right) D(Y_\alpha^{(j)} - K^{(ij)}) + \left(\frac{1 - M^{(ij)}}{2}\right) \left(1 - D(Y_\alpha^{(j)} - K^{(ij)})\right). \quad (\text{S5.56})
\end{aligned}$$

Here  $D(Y_\alpha^{(j)} - K^{(ij)}; 0, \sigma K^{(ij)})$  is the cumulative distribution function of the normal distribution with mean 0 and standard deviation  $\sigma K^{(ij)}$  evaluated at  $Y_\alpha^{(j)} - K^{(ij)}$ . Because of the AND-logic we impose, and because the noise terms are independent for each  $(i, j)$ , the probabilities for the final state can be written as

$$P_\alpha(\beta^{(i)}) \equiv P(\beta^{(i)} | \alpha) = \begin{cases} 1 - P(\mathbf{g}_\alpha^{(i1)} = 1)P(\mathbf{g}_\alpha^{(i2)} = 1) & \beta^{(i)} = 0 \\ P(\mathbf{g}_\alpha^{(i1)} = 1)P(\mathbf{g}_\alpha^{(i2)} = 1) & \beta^{(i)} = 1 \end{cases} \quad (\text{S5.57})$$

$$= \beta^{(i)} \prod_j P(\mathbf{g}_\alpha^{(ij)}) + (1 - \beta^{(i)}) \left(1 - \prod_j P(\mathbf{g}_\alpha^{(ij)})\right). \quad (\text{S5.58})$$

By inserting Equation S5.56 into Equation S5.58 we obtain transition probabilities at the single-cell level in terms of the sensed concentrations  $Y_\alpha$ , thresholds  $K^{(ij)}$  and noise levels  $\sigma$  (for a given transition  $\alpha \rightarrow \beta$ ). In order for the entire wave to propagate, the transitions must be satisfied for each gene  $i$  and each of the cells in the system. Let  $n_\alpha$  be the number of cells of type  $\alpha$  in our system. For a single, straight plane wave, we have  $n_\alpha = \sqrt{N}$  for  $\alpha = E_F, F, M, B, E_B$  and  $n_\alpha = N - 5\sqrt{N}$  for  $\alpha = E$ . We can then write the probability that the wave survives for one time step as

$$P_{survival} = \prod_\alpha P(\alpha \rightarrow \beta)^{n_\alpha} = \prod_\alpha \left( \prod_{i=1}^L P_\alpha(\beta^{(i)}) \right)^{n_\alpha}. \quad (\text{S5.59})$$

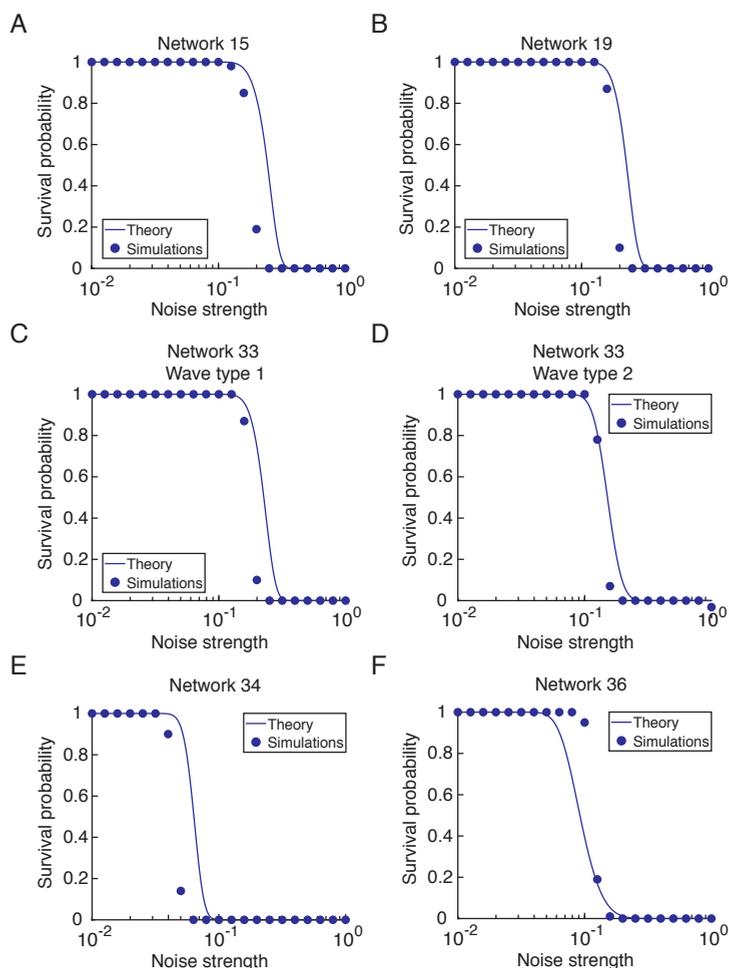
The probability that the wave survives for  $t$  time steps is then  $(P_{survival})^t$ .

**Accuracy of computed survival probability** The accuracy of the computed survival probability mostly depends on the validity of the nearest-neighbor approximation. Recall that we estimate the signal molecule concentration from all cells beyond nearest neighbors through a mean-field approximation term  $Y_{MF}$  (Equation S5.39). This ap-

proximation is the reason why the computed TW propagation conditions are not exact, and as a result the survival probabilities are also not exact. However, this only leads to significant discrepancies when either (1) the interaction strength is very high, i.e. when cells interact strongly with each other ( $a_0$  small), or (2) when the original parameter set is close to the boundary of the region where TWs can propagate. In the second case, the problem arises when the calculated boundary of the ‘TW propagation phase’ is not accurate. In this case, the noise strength required to perturb the system beyond the boundary cannot be accurately estimated, which in turn leads to the survival probability being inaccurately computed. When the system is far away from the boundary, these deviations are proportionally smaller and as a result the order of magnitude estimation of the required noise strength to perturb the wave is more accurate.

**Application to other complex elements** Similar arguments can in principle be derived to study the effect of disordered cell positions and cell motility, while for finite Hill coefficient at this point we lack an analytic framework for wave propagation. In essence, when only the positions of the cells are altered, this reflects only in the interaction terms  $f^{(i)}(r_{ij})$  which are functions of the distance between cells  $r_{ij}$ . At first approximation, for small deviations only the nearest neighbor terms  $f_{nn}^{(i)} = f^{(i)}(a_0)$  are affected. Hence one can estimate the effect of spatial rearrangement of the cells through the effect on the nearest-neighbor interaction strengths  $f_{nn}^{(i)}$ , which now typically become different for individual cells if the rearrangement process is stochastic. These in turn can be directly computed from distributions of nearest-neighbor distances obtained from the stochastic rearrangement process.

However, in practice, even for simple Brownian motion of the cells the expressions for distributions of nearest-neighbor distances yield unwieldy mathematical expressions in terms of special functions such as Bessel functions. Furthermore, while the effects are straightforward to estimate for a pair of moving cells, it is not trivial whether the pairwise calculation can be extended to a full lattice. As such, we have not attempted to fully derive analytic results regarding the effect of moving cell positions on traveling wave propagation, but merely want to point out that such a calculation is in principle possible.



**Figure S5.13: Persistence of traveling waves in the extended model with stochasticity: theory and simulations (see Section S5.5.4) (Related to Figure 5.6)** The subfigures have the following fixed parameters:  $N = 225$ ,  $a_0 = 1.5$ ,  $r_{cell} = 0.2$ ,  $C_{ON} = (500 \ 500)$ , and (A)  $\begin{pmatrix} 50 & 300 \\ 300 & 0 \end{pmatrix}$ , (B)  $\begin{pmatrix} 50 & 300 \\ 300 & 0 \end{pmatrix}$ , (C)  $\begin{pmatrix} 40 & 500 \\ 300 & 50 \end{pmatrix}$ , (D)  $\begin{pmatrix} 50 & 500 \\ 300 & 70 \end{pmatrix}$ , (E)  $\begin{pmatrix} 900 & 300 \\ 400 & 50 \end{pmatrix}$ , (F)  $\begin{pmatrix} 900 & 400 \\ 300 & 50 \end{pmatrix}$ . (A-F) We used the analytic framework for studying wave propagation (Figure 5.4) to understand how adding noise affects wave propagation. We introduced noise by stochastically varying the threshold concentrations (i.e.,  $K^{(ij)}$ ) and defined “noise strength” that quantifies the typical fluctuations in  $K^{(ij)}$  (see Section S5.4.3). The plots here show “survival probability” – the probability that a rectilinear wave, after forming, persists (keeps on traveling) for at least one full period – as a function of the noise strength for all six cellular dialogues that we identified as capable of forming dynamic spatial patterns (Figure 5.4D). Solid curves in the plots are from our analytic framework, which had no fitting parameters. The data points are from simulations, with 100 iterations of simulations performed for each data point. These results are for fixed sets of parameter values which we chose to lie well within the region of the parameter space where one obtains persistent waves (see, for example, Figure S5.9B for these regions).

# III

## EPILOGUE

# 6

## DISCUSSION

In this thesis, we have examined the collective dynamics of populations of communicating cells by building mathematical models, developing theoretical concepts and metrics and analyzing their behavior through numerical simulations and analytic calculations. In the following, we state two general conclusions, which relate to the various results presented throughout the thesis. We then critically examine the biological realism of our models and discuss different experimental strategies for validating our results. Finally, we offer a list of suggestions for further research.

### 6.1. KEY LESSONS LEARNED

#### CONCLUSION 1: MOLECULAR-SCALE PROPERTIES DICTATE EMERGENT POPULATION-LEVEL BEHAVIOR

Throughout this thesis, we have aimed to relate microscopic parameters to emergent population-scale dynamics, for a generic class of multicellular systems consisting of secrete-and-sense cells. This effort mirrors the approach taken to construct genotype-phenotype maps in evolutionary biology. We find that microscopic features of cell-cell interactions control macroscopic dynamics by constraining the range of possible population-level dynamics. This is evident in both the role of the underlying gene regulatory network as well as that of the molecular parameters specifying processes, such as diffusion of the signaling molecules, secretion and response of the cells. For instance, we have characterized the conditions required to propagate traveling waves and found that they are only

satisfied by a subset of two-gene networks and their corresponding regions in parameter space. Thus, by tuning these molecular-scale properties, we can to a great extent control the collective behavior of a population of interacting cells.

Experimentally, this idea was demonstrated in engineered yeast strains that communicate by secreting and sensing the same signaling molecule [Youk & Lim, 2014a]. Here, the authors showed that the social behavior of quorum sensing cells can be tuned by modulating the sensing and secretion pathways through chemical inducers. While the experimental study was done in bulk, in our models we can simulate a spatially distributed collection of cells. Earlier work already showed that the collective behavior of such a system can be characterized by different “phases”, whose boundaries are determined by parameters underlying the response of the cells and the effective cell-to-cell distance [Maire & Youk, 2015a]. We have expanded upon this idea by calculating an effective “pseudo-energy” landscape, whose shape is dictated by these molecular parameters, and which constrains the dynamics of our system (Chapter 4). In this way, we obtained a visual representation linking the molecular parameters to the overall “phase” of the system.

With multiple signaling molecules, the primary determinant for the types of patterns that the system can generate is the underlying gene network topology (Chapter 5). We considered simple topologies, where the different molecules interact at the genetic level by regulating each other's production levels, and showed that distinct architectures can generate distinct patterns. For instance, we have shown that there are classes of gene regulatory networks that are incapable of forming any dynamic pattern, regardless of the tuned molecular parameters (Chapter 5 – Figure 5.3). Furthermore, once the appropriate network topology is chosen, the parameters still need to be tuned accordingly to generate the required patterns. For instance, to generate dynamic spatial patterns such as traveling waves, we found that the strength of the positive self-interaction(s) need to be strong (Chapter 5 – Figures 5.4F and S5.9). Our analytic framework shows that it is possible to deduce general constraints on molecular parameters that determine whether such patterns can form and propagate (Chapter 5 – Section S5.5.3 and Figure 5.4). Furthermore, we have also derived general constraints on the system dynamics that arise from molecular parameters, which can be stated in terms of state diagrams (Chapter 5 – Section S5.5.2 and Figure S5.7). Together, these results show that it is possible to derive constraints on the macroscopic behavior of our system directly from microscopic parameters.

## CONCLUSION 2: STATISTICAL PHYSICS CONCEPTS ELUCIDATE SYSTEMS-LEVEL PROPERTIES OF MULTICELLULAR SYSTEMS

A second recurrent theme throughout this work has been the use of tools from statistical physics to study the collective behavior of interacting cells. In recent years, statistical physics has branched out from its original focus on studying molecules to studying firing neurons, bird flocks and the immune system [Mora & Bialek, 2011]. These studies typically find an exact correspondence between predictions of a statistical physics model, such as the Ising model, and biological data. For instance, the spike trains of individual neurons can be represented as binary sequences representing whether a neuron is spiking or not. By studying correlations between these spiking events in different neurons, researchers proposed that pairwise interactions — such as captured in the Ising model — are sufficient for understanding these correlations [Schneidman et al., 2006]. A more fundamental question arising from these approaches is whether biological systems are poised at criticality [Mora & Bialek, 2011]. While this may seem like an abstract concern, if a biological system is indeed close to a critical point, this would enable us to study it using the arsenal of statistical physics knowledge on critical phenomena.

## 6

While our work has drawn inspiration from these approaches, we did not directly apply statistical physics to biology, due to basic incompatibilities with our model. For instance, our system lacks natural definitions of key quantities such as temperature and energy. Although we have postulated analogous quantities such as the “pseudo-energy” (Chapter 4) and “multicellular entropy” [Maire & Youk, 2015a], these are strictly speaking not thermodynamic quantities. One way to resolve these incompatibilities is by directly employing Ising models to model pattern formation [Hillenbrand et al., 2016]. This has the advantage of starting from a well-studied physics model, but requires assuming phenomenological quantities (e.g. coupling terms not directly derived from biophysics). In contrast, all our model elements arise directly from modeled biophysical and biochemical processes, but this means that it no longer directly corresponds to a physics-type spin model (more on this later).

In Chapter 3.2, we studied the transition between a (disordered) “autonomous phase” and an (ordered) “collective phase” in a population of cells with continuous states. While this phenomenon is simple and perhaps unremarkable, we motivate our study by referring to a highly popular model in statistical physics that exhibits a similar phenomenon. Indeed, the Vicsek model — arguably the quintessential active matter model — is also an agent-based model that shows a density-dependent transition between an ordered and a disordered phase [Vicsek et al., 1995]. Although the Vicsek model is a highly simplified representation of flocking birds, its tremendous popularity can perhaps be attributed to the success of applying statistical physics to understand the density-dependent transi-

tion as a second-order phase transition [Ginelli, 2016].

In our approach, we have also sought to characterize the density-dependent transition between “autonomy” and “collectiveness” as a phase transition. This is evident from the metrics we developed for characterizing this transition, which all show rather sharp transitions between the two phases. These metrics should be regarded the equivalent of “order parameters” to characterize this transition. In addition, we have also used mean-field theory to explain observations for weakly interacting cells, which is a commonly used strategy in statistical physics. Although we did not develop a rigorous method to characterize this transition as a phase transition, our attempts have identified metrics to characterize statistical properties of this transition which are also applicable to other systems with similar phenomena.

In Chapter 4, we took the statistical-physics analogy further by exploiting the analogy between our model and the Ising model. Here, the correspondence is very strong, but we should give a note of caution to avoid readers from confusing our model as simply an instance of the Ising model. As mentioned, our model does not have natural definitions of thermodynamic quantities such as temperature and energy. Moreover, the dynamics of our model arises from a cellular automaton with *deterministic* rules. This cannot directly correspond to the Ising model or any spin model. At finite temperature there is necessarily noise in the system (e.g. as modeled by Glauber dynamics), whereas at zero temperature the ground state cannot have spatial structure, since it is trivial (e.g. homogeneous for ferromagnetic interactions). This is in contrast to *probabilistic* cellular automata, which in certain cases can be directly mapped onto Ising models [Merle et al., 2019].

As such, our framework is only analogous to statistical physics frameworks, but not identical to any existing statistical physics model. Nevertheless, let us discuss the analogy in a bit more detail. Our macrostate description in terms of  $p$  and  $I$  is analogous to an Ising macrostate with fixed energy and average magnetization. In fact, the average gene expression has the same mathematical form as the average magnetization. Furthermore, the spatial index  $I$  is to some extent comparable to a correlation length, except in the cases where the lattice becomes (almost) homogeneous. Then our spatial index tends to zero (see Section S4.5.1), whereas the correlation length would diverge. Our “pseudo-energy landscape” with its stochastic equation of motion mirrors free energy minimization for systems in thermodynamic equilibrium. Yet our equilibrium states are not located at free energy minima (even local ones), because our energy landscape does not contain any minima except at the extremes where the system is homogeneous (see Section S4.5.6). This comes from the redundancy of our macroscopic description — many

states may correspond to the same macrostate. As a result, equilibrium states and non-equilibrium states (as determined by the cellular automaton's rules) may correspond to the same macrostate (have the same value of  $p$  and  $I$ ). Therefore, we introduced another concept — the equilibrium probability or intuitively the landscape's "stickiness" (Section S4.5.5) — to obtain an approximate picture of the locations of equilibrium states on the energy landscape.

## 6.2. RELATION TO OTHER MATHEMATICAL MODELS

In this section, we briefly discuss the relation between the multicellular model studied in this thesis and two broad classes of models that show similar self-organization phenomena: cellular automata and coupled map lattices. We will sketch how our results relate to some well-established concepts and features of these systems.

### CELLULAR AUTOMATA

As a discrete-time, discrete-space and discrete-state model, our model in its original form (Chapter 3.1) is an instance of a cellular automaton (CA). However, note that a narrower definition of CA would include only models where the interactions are local, i.e. the update rule depends only on the nearest neighbors of a cell. In contrast, our model the cells are globally coupled (Eq. 3.7), i.e. each cell has a non-zero interaction with all cells in the system. However, at large distances the interaction function decays exponentially with distance to a cell (Eq. 3.1). Therefore, if the interaction range  $\lambda$  is small compared to the inter-cell distance  $a_0$ , the interactions are approximately local.

Much of what is known about cellular automata stems from studies of the elementary CA [Wolfram, 1984; Wolfram, 2002]. Our model is not an elementary CA, because of the mentioned global coupling and because it is 2D. Furthermore, while the elementary CA form a discrete set of model instances, our system constitutes a family of models parametrized by continuous variables. As such, each instance of our model with a fixed set of parameters could in principle correspond to a different elementary CAs under this approximation.

Recall that the elementary CAs have been classified into four classes:

1. Class 1: nearly all configurations evolve towards a uniform steady state.
2. Class 2: nearly all configurations evolve towards stationary or periodic steady states.
3. Class 3: nearly all configurations lead to aperiodic ("chaotic") patterns.
4. Class 4: nearly all configurations lead to complex patterns that can interact in in-

interesting ways, rendering the system suitable for universal computation (i.e. a Turing machine).

Our system can be identified with different classes, depending on how the parameters are chosen. For one gene, the “phenotype diagrams” [Maire & Youk, 2015a] can be used to identify Class 1 behavior, which would only occur in the “all ON” and “all OFF” phases. As a single gene does not induce chaos or complex patterns, in all other phases the system falls into Class 2. With multiple genes, the state diagrams (Chapter S5.5.2) directly identify when the system is in Class 1. In this case the state diagram would contain a single steady state and no loops. Class 3 and 4 only occur for the network topologies capable of generating dynamic spatial patterns (Fig. 5.3). Intuitively, a system capable of forming traveling and other propagating waves would be Class 4, as they have the capacity to interact in complex ways, annihilate and generate new waves. However, it is not known whether our system is Turing complete.

Wolfram has also developed “statistical mechanics” of CA [Wolfram, 1983]. This framework does not literally correspond to statistical *thermodynamics* with its Boltzmann distribution and thermodynamic ensembles, which is also the case for our framework (as discussed in Conclusion 2, Chapter 6). Rather, it is a set of tools to characterize statistical properties of the configurations and dynamics of CA. Some of our quantitative metrics are borrowed from or have a direct counterpart to Wolfram’s framework. For instance, whereas Wolfram uses two-point correlation functions from physics, we used Moran’s I as a way of characterizing spatial correlations. Both quantify spatial correlations, but from correlation functions we obtain a length scale (the correlation length). However, defining a correlation length is challenging when the correlation function does not decay exponentially, as is the case in our system. The Hamming distance is also borrowed from this work, where Wolfram uses this to show that in the chaotic regime, initial states differing by a single cell can diverge from each other exponentially. Furthermore, he proposes that self-organization can be understood as an entropy-reduction process, whereby the Shannon entropy of the distribution of accessible states goes to zero. This principle should also hold for our system, but the Shannon entropy is hard to compute due to the large number of possible states.

### COUPLED MAP LATTICES

A similar comparison can be made between our model and coupled map lattices (CMLs). When the response function has a finite Hill coefficient (Chapter 3.2), the state variable, i.e. the gene expression level of a cell, becomes continuous, rendering our system a CML. Kaneko has proposed different types of CML behavior, which he dubbed “universality classes” [Kaneko, 1989; Kaneko, 1992]. Examples include “frozen random patterns”,

where the system forms a stationary or oscillatory temporal pattern but a random spatial pattern; “pattern selection”, where domains of a certain size become dominant over time; traveling waves; and “fully developed spatiotemporal chaos”.

At a qualitative level, we recognize a few elements of these phases in our model, but there seems to be no exact correspondence between the logistic CML phases and our model’s phases. For instance, the “frozen patterns” would correspond to the case where cells are autonomous. This would essentially “freeze” a random pattern of independent cells that either remain stationary or oscillate, depending on the underlying gene circuit and parameters. However, we have not observed any clear examples of “pattern selection”, where the pattern consists of domains of a limited size. We have also not found a clear correspondence to any of the other logistic CML phases identified by Kaneko [Kaneko, 1993].

In terms of analysis, we can use a variety of analytical tools to characterize CMLs. Kaneko proposed that the different regimes he defines have distinct signatures in terms of a number of metrics [Kaneko, 1989]. In particular, he showed that the spatial and temporal power spectra (obtain from Fourier transforms) look different in the various phases of the logistic CML. The same is true in our multicellular model, where the spatial spectra can carry distinct signatures for distinct phases<sup>1</sup>. This is for instance the case for (straight) travelling waves, for which the spatial power spectrum shows distinct peaks for the modes corresponding to the orientation of the wave. As such, we should be able to quantify the degree of “waveness” using power spectra. This idea was further explored by defining a “wave score”, but this quantity was found to be not applicable to all waves [Grundel, 2019].

## 6

### 6.3. BIOLOGICAL REALISM OF OUR MODELS

We have aimed to construct minimal models that incorporate biologically realistic features such as cell signaling and response in the simplest possible ways. As such, we necessarily omit features that may be present in some multicellular systems, especially in the development of complex organisms. We have already motivated modeling choices such as bistable cells arising from ultrasensitivity (Chapter 3.1) and separation of timescales between molecule and gene expression dynamics (Chapter 4.3). We have also mentioned active transport, clustering and receptor endocytosis as additional contributing processes (Chapter 4.3). Here we elaborate on a number of such effects that we have not taken into account and suggest how we may expand our framework to include these effects. Adding these additional features may or may not change the results obtained from

<sup>1</sup>Detailed results of these and other efforts to quantitatively characterize patterns are presented in the master thesis of Douwe Grundel [Grundel, 2019].

our modeling efforts. Either way, we have already shown that none of these features are necessary for generating complex patterns. As such, adding these features should only be a priority in cases where the minimal model is clearly shown to be insufficient.

First of all, there are several other sources of cell-cell interactions that may be significant in real tissues, such as mechanical interactions and contact-mediated signaling (reviewed in Chapter 2). Adding contact-mediated signaling is straightforward on a lattice model, but it is less obvious how to model its downstream effects and how it interacts with the secrete-and-sense mechanism. Mechanical interactions play an important role in a variety of morphogenetic processes and especially the interplay between mechanical and chemical signaling can be significant in development [Hannezo & Heisenberg, 2019]. However, for interacting unicellular organisms such as social amoebae, mechanical effects would be less significant. Including mechanical interactions in the current model would likely require major changes to the model. Our assumption of non-touching spherical cells would need to be heavily modified, and with that also the description of chemical signaling (which relies on this geometry when solving the underlying reaction-diffusion equation). We would then likely need to model the cells' shapes and geometry-dependent mechanical interactions explicitly. One possibility to circumvent this is by modeling cells as connected springs, which appears sufficiently realistic to explain experimentally observed dynamic patterns [Aoki et al., 2017].

We have used simple equations for describing regulation at the genetic level, involving a single response function (either step-like or sigmoidal) and direct genetic interactions in the case of multiple signaling molecules. Real biological networks are complex and involve far more components. However, as theorists we must make simplifications to keep models tractable, and choosing the correct simplification may pose a challenge. When it comes to the complexity of gene networks, we note that make simplifications by combining elements of e.g. the signaling response network, and describe only its effective outcome [Tyson et al., 2003]. This is also the main idea behind modeling ultrasensitive response, which may arise from multistep processes, as single step functions or sigmoidal functions (discussed in Section 3.1). Nevertheless, we should acknowledge that the choices we made are not unique. There are still many possible response functions depending on the implicit topology of the signaling network, and future work may expand to include these other possibilities [Tyson et al., 2003].

Furthermore, while our main model elements are based on underlying physics and chemistry, the elements of our model expansions are mostly phenomenological (Section S5.4.3). The reason for this is that the modeled processes — in particular stochasticity and cell motility — are highly complex, and modeling these is not the main focus of our work.

As such, we have employed a highly simplified, phenomenological description of noise and cell motility in the system. In the case of cell motility, we have already mentioned biologically realistic alternative descriptions (Chapter S5.4.3). As for noise, we should mention that stochasticity enters not only at the cell sensing and gene regulatory level, but also in the response times of cells, arising from the underlying chemical kinetics of often multi-step reaction pathways [Thurley et al., 2018]. Hence, the fact that we update all cells synchronously in our cellular automaton may not always correspond to realistic situations, but is still a good approximation for when the response time distributions are narrow. Alternative updating schemes for the cell states, e.g. based on the stochastic simulation algorithm or any of its adaptations [Cao et al., 2005], could alleviate this concern.

Finally, while in our idealized setting cells are identical and are always actively secreting and sensing, real tissues tend to be more heterogeneous, even when all cells are of the same type. As such, the actual interactions between cells, as determined by correlations in their dynamic activity, tends to have the structure of complex networks with a high degree of clustering, rather than that of homogeneous networks [Gosak et al., 2018]. One example is found in the pancreas, where  $\beta$ -cells communicate through gap junctions and paracrine signal. Correlations in calcium spiking of these  $\beta$ -cells (a proxy for insulin release) show that spiking activity is coordinated in a highly nontrivial fashion, with distinct clusters of cells that have highly coordinated activity within the same cluster, but not between different clusters [Gosak et al., 2018]. This suggests that tissues could have more complex interaction networks than the homogeneous systems we have considered. To take this into account, we could alter our model by taking a random graph approach, where we vary the connectivity between cells by randomly assigning links.

## 6.4. EXPERIMENTAL VALIDATION

The ultimate test of the biological realism of our models can only be obtained through direct comparison with experimental data. We have already mentioned optogenetics (Chapter 4.3) and synthetic gene circuits (Chapter 5.3) as possible methods for directly testing model predictions. Here we discuss in more depth various strategies for relating our findings to experiments. As our model has minimal assumptions, we envision applying its results to various types of systems. Here, we shall distinguish between developmental biological systems, synthetic biological systems and non-biological systems. The focus is mainly on dynamic patterns, but much of the discussion will be general and applicable to any type of pattern formation.

### 6.4.1. DEVELOPMENTAL BIOLOGY SYSTEMS

In developmental biology systems, the constituents and interactions — e.g. cell types, genetic networks, diffusion constants — are in principle fixed. Experimentalists can adjust these parameters to a limited extent by changing the environment of the system, e.g. through the addition of chemical components, or by performing genetic perturbations. Moreover, a multitude of possible interactions may be at play simultaneously, many of which may be unknown. Finally, as there are many mechanisms of generating similar phenomena, testing ideas about proposed mechanisms in natural systems is typically challenging. Altogether, this makes directly testing theoretical hypotheses in naturally occurring developmental systems more challenging than in synthetic systems.

Nevertheless, a number of recent studies have tackled the challenge of explaining various dynamic patterns in developmental systems through theoretical models [Lubensky et al., 2011; Cotterell and Sharpe, 2015; Jörg et al., 2019]. The general approach in these studies is to start with gene networks and cellular interactions as known from experiments, and then to model the spatiotemporal dynamics of relevant genes or processes as resulting from these interactions. The patterns that are produced are then compared with experimental results. If mutants with deviating phenotypes are available, these can consolidate proposed theoretical models if the data matches. This strategy is feasible also for verifying our models. Since many different mechanisms can lead to the same pattern, probing the dynamics of pattern formation and examining how a system responds to perturbations is required to claim validity of a theoretical model.

In terms of gene network architectures, our results are promising when compared with these and other theoretical and experimental findings. The above studies reproduced traveling waves with gene networks and cell-cell interactions which are more complex (i.e. larger and with more links) than the ones considered in our computational search. Nevertheless, their general topological features qualitatively match with our results. In all of these works, the underlying interaction network consists of interlocked positive and negative feedback loops, where the positive feedback is typically direct (e.g. without explicit intermediate mediators), whereas the negative feedback is mediated through interaction with a secondary component. This corresponds precisely to the “activation-inhibition” cellular dialogues capable of generating dynamic spatial patterns, e.g. network 15 and its derivatives (Figure 5.1).

With regard to the FitzHugh-Nagumo (FHN) model, a prototype for describing excitable phenomena, these results should not be surprising [Gelens et al., 2014]. The main qualitative feature shared with the FHN model is the combination of a fast activator and a slow inhibitor, leading to transient excitation spikes that form traveling waves in a spatially

distributed system. However, despite the success of FHN model in explaining diverse biological phenomena [Sgro et al. 2015; Hubaud et al., 2017], it remains a phenomenological model whose constituents can typically not be mapped directly to biophysical quantities. Conversely, in the bottom-up modeling approach that we and others took, a qualitatively similar mechanism arises from biophysically realistic assumptions.

Furthermore, the networks we identified as the ones generating dynamic spatial patterns also correspond precisely to the networks generating Turing patterns in reaction-diffusion systems [Scholes et al., 2019]. Why does our model not seem to generate Turing patterns then? The main assumption that distinguishes our model from reaction-diffusion systems models is that we impose a separation of time-scales between the signaling molecule dynamics and the gene expression dynamics. Delays in gene expression response time could well be responsible for generating different phenomena. This is exemplified by gene circuits with a single negative feedback. When the response is immediate, this typically leads to dampening of oscillations, such that all dynamics leads to a single stable fixed point [Alon, 2006] — this situation corresponds to homeostasis. However, if there is a delay in response time, the same circuit leads to stable, never-ending oscillations. Likewise, it is possible that traveling waves in our model are a result of the slow cellular response compared to signaling molecule dynamics. Another possibility is that static patterns with a well-defined wavelength (i.e. Turing patterns) are possible, but their wavelengths are larger than the typical simulated system size (i.e. hundreds of cells). Simulating larger systems and characterizing correlation lengths would be a way to tell whether this is the case.

## 6

#### 6.4.2. SYNTHETIC BIOLOGICAL SYSTEMS

Synthetic biology provides another promising avenue for studying multicellular pattern formation. Engineering patterns in controlled setups can provide new insights about patterning mechanisms and allow for direct comparison with computational models. As Richard Feynman noted, “What I cannot create, I do not understand”. Typically, one starts from well-known mechanisms of pattern formation and aims to reconstitute patterns in populations of cells with engineered genetic circuits, possibly under the influence of external effects such as imposed morphogen gradients. These studies have successfully reconstituted a variety of patterns, including stripes and salt-and-pepper patterns, and employ a various patterning mechanisms including morphogens, cell signaling, phase separation and mechanical forces (reviewed in [Scholes and Isalan, 2017; Santos-Moreno and Schaerli, 2018]).

However, studies of dynamic patterns in synthetic systems are scarce, and have mostly been limited to synchronized oscillations. An early milestone paper in engineering syn-

thetic patterns shows how coupling cell signaling to cell motility results in a pattern of concentric rings in an expanding population [Liu et al., 2011]. Another study engineered a quorum-sensing module which is capable of generating synchronized oscillations [Danino et al., 2010]. Interestingly, this system is also able to produce traveling waves, and has a gene circuit matching one of the circuits identified in our computational screen as capable of forming dynamic spatial patterns. However, this wave should be understood as arising from scale limitations in oscillation synchronization, i.e. the fact that the oscillations synchronize only locally but cannot globally couple the whole system [Santos-Moreno and Schaerli, 2018]. In a large enough population, this could lead to phase waves or kinematic waves, as a result of incomplete synchronization of oscillations. Thus, despite having the same network topology and generating the same phenomena, it remains questionable whether our waves arise from a similar mechanism as in the Danino *et al.* paper.

A better understanding of dynamic patterns in these systems could be obtained from more controlled experiments, which would also enable better quantitative comparison with our modeling results. To start with, one would need more precise control over the spatial positioning of cells, beyond what can be obtained by plating cells on a petri dish. This can be achieved through both micropatterning techniques that simulate native physiological conditions [Théry, 2010], as well as microfluidics setups, which enable spatial confinement and constant medium refreshment (also discussed in the next section). Furthermore, effects of cell proliferation and death should play a minimal role during the time course of an experiment, which can be achieved by inhibiting cell division using standard mitotic inhibitors. Finally, we would need to image the gene expression dynamics of our system over sufficiently long times to study dynamic patterns.

Furthermore, synthetic engineering could allow for a comprehensive effort to screen a large number of different quorum sensing circuits and study population-level dynamics for populations with each of these circuits. This would first require constructing strains with these different signaling motifs and with tunable genetic interactions and environmental conditions. For instance, interaction strengths between modular components could be tuned through inducible promoters [Youk & Lim, 2014a]. Subsequently, one would study the spatial dynamics of these systems using standard microscopy techniques, preferably in combination with the features mentioned in the last paragraph. As screening a large set of gene circuits and parameters ranges may be unrealistic when done manually, one could potentially benefit from high-throughput imaging assays [Pep-perkok & Ellenberg, 2006].

Pattern formation represents only the first step in development, and is typically fol-

lowed by differentiation and morphogenesis. Engineering spatial patterns is thus a part of wider attempt to utilize synthetic biology to understand development (reviewed in [Davies, 2017; Ebrahimkhani & Ebisuya, 2019]). Whereas we have only dealt with pattern formation in this thesis, future theoretical work could aim to study development in models integrating pattern formation with differentiation and morphogenesis. Theoretical approaches combining chemical patterning with mechanical interactions are a promising step in this direction [Brinkmann et al., 2018; Recho et al., 2019]. Validation of these theoretical models requires further connection with synthetic biology attempts to recreate developmental processes *in vitro*.

### 6.4.3. NON-BIOLOGICAL SYNTHETIC SYSTEMS

Finally, a third promising avenue is to explore our ideas in non-biological synthetic systems, most notably those based on DNA nanotechnology. Historically, this field has been associated with structural DNA nanotechnology – e.g. how to form self-assembling structures, crystals, DNA origami, and so on. However, in the last two decades or so, the field of dynamic DNA nanotechnology – how to engineer DNA-based systems that perform dynamic functions – has flourished and now enables the construction of large chemical reaction cascades, logic circuits, complex dynamical systems and neural networks [Zhang & Seeman, 2011]. Specifically, many of these dynamic functions are based on *toehold-mediated strand displacement*, whereby a single stranded input strand reacts with a multi-stranded DNA complex, thereby displacing another bound single strand, which gets released from the complex as the output. This mechanism enables the equivalent of a chemical reaction in a controlled, enzyme-free process, and can be scaled up for the engineering of more complex reaction networks.

In the context of our work, these approaches could be utilized to construct synthetic communication systems, which could be used as an abiotic model to study communicating biological cells. A major advantage of using such non-living synthetic systems is that they potentially confer a much higher degree of experimental control. Negative side-effects such as interactions with endogenous pathways and molecular crowding would be absent or reduced. Furthermore, such systems allow for easier tuning of molecular parameters, whereas *in vivo* these are typically set by the characteristics of the host cells [Dubuc et al., 2019].

A key challenge in the construction of synthetic communication networks to study spatial patterns is enabling spatial control, which would typically require some form of compartmentalization. Various microfluidic technologies exist for studying artificial gene networks in compartmentalized devices, including flow reactors, microdroplets, liposomes and coacervates [Dubuc et al., 2019]. Other approaches localize DNA circuits on

a surface, for instance through interactions with DNA origami [Chatterjee et al., 2017]. These technologies enable the study of cell signaling beyond well-mixed systems, or even reaction-diffusion systems, which are typically in a continuous medium without (artificial) cells [Loose et al., 2008]. The combination of compartmentalization (e.g. in liposomes or other protocells) and localization (e.g. through microfluidics) enables the study of spatial patterns in realistic settings resembling natural systems.

These approaches have yielded synthetic, cell-free systems that reproduce a wide range of biological phenomena. An early study reconstructed the *Drosophila* gap gene network using DNA-coated magnetic beads fixed in an artificial chamber [Isalan et al., 2005]. Another study constructed “ring oscillators” consisting of a series of molecules that each repress the next and exhibit oscillations [Niederholtmeyer et al., 2015]. Synchronization or entrainment of coupled individual oscillators was studied in a 1D microfluidic device consisting of DNA compartments [Tayar et al., 2017]. Finally, a recent study reports a molecular communication platform based on protein-based synthetic protocells, that can have orthogonal communication channels [Joesaar et al., 2019]. Combined with a microfluidic protocell trap array and confocal imaging, this setup can be used to study the spatiotemporal dynamics of dozens of communicating protocells. Specifically, the authors have demonstrated signal amplification, signaling cascades, negative feedback, Boolean logic in this system. This work is closest to what we have studied theoretically, and an extended version of this platform could well be used to test our hypotheses.

To make this connection between theory and experiment, two main challenges need to be overcome. First, one would need to engineer more complex circuit structures than the ones currently published. A number of the mentioned studies already feature multiple signaling components, but these are coupled in a simple fashion (e.g. in a linear cascade). Realizing the circuits of our model would require integrating multiple communication channels, feedback loops and logic gates into a single (proto)cell. Furthermore, current studies focus on particular realizations of cell signaling motifs to demonstrate proof of concepts, but a more comprehensive picture is lacking. Engineering large libraries of different circuits with tunable properties and studying their spatiotemporal dynamics would be ideal in this case, but may be unrealistic at the present moment. However, if successful, such an approach would enable a more comprehensive understanding of the versatility of multicellular signaling systems.

## 6.5. DIRECTIONS FOR FUTURE RESEARCH

Given the previously stated conclusions and experimental relevance, what are some concrete steps to take next, if we were to continue with this project? In this final section,

we discuss possible directions for future research. The aims of these suggested projects are manifold. To begin with, they serve to directly test the current theoretical framework and improve upon it if necessarily (points 1-2 below). This direction requires direct collaboration with experimentalists, but also complementary modeling efforts. These could in return guide further development of our modeling framework to incorporate a wider range of biologically realistic settings. Secondly, we should aim to extrapolate our framework and findings to more general settings, to show the wider applicability of our work (points 3-5 below). This involves comparing our findings to related modeling approaches (point 3), as well as deriving generalized mathematical descriptions beyond the special cases we considered (point 4), which includes extrapolating specific results to more general settings (point 5). Thirdly, we should explore specific features of our models in more detail, as they are currently incompletely understood (points 1, 5 and 6). This requires the development of new theoretical tools and concepts. Finally, we should aim to place our results in the wider context of literature from a variety of fields (points 7).

### **1. Expand current modeling efforts to include more complex settings and by incorporating additional features relating to real biological systems (as discussed earlier)**

The first point relates to the biological realism of our models, as discussed in Section 6.3 earlier in this chapter. Priorities in this direction should include modeling more complex networks, multiple cell types and cell-cell heterogeneity. Expanding our framework to include more complex networks is straightforward, but note that the number of distinct networks grows rapidly with the number of nodes, i.e. the number of molecule types. Hence extending to more molecules requires not only considerably more computational power, as well as automated procedures for analyzing large sets of simulations (for instance by defining a “wave score”, as discussed under point 6). Secondly, real tissues are often composed of multiple cell types, each of which may have a different cellular dialogue and molecular parameters. A formal extension to multiple cell types for cells with single signaling molecule is given in the supplementary section of [Olimpio et al., 2018] (omitted from this thesis as it is work by E. Olimpio), but the full range of phenomena remains to be explored. Thirdly, populations of cells are never fully homogeneous, even when they are genetically identical [Ackermann, 2015]. This calls for models to take into account cell-to-cell variability. Such variability could be introduced simply by taking random molecular parameters for each cell, preferably drawn from experimentally characterized distributions. Related to this point, we should also explore more realistic descriptions of stochasticity. Rather than taking a single phenomenological parameter,

we could try to derive more exact descriptions of noise arising directly from biophysical sources, such as fundamental limits in sensing accuracy [Berg & Purcell, 1974] and stochastic gene expression. Interestingly, in reaction-diffusion systems the addition of noise produces so-called stochastic Turing patterns [Biancalani et al., 2010], which are similar to normal Turing patterns and have the same wavelengths, but have a wider parameter space under which they can self-organize. They have since then been observed experimentally in a synthetic bacterial population [Karig et al., 2018]. In our model, we have so far observed noise-induced ordering in both static pattern formation (Chapter 4 – Section 4.2.7) and dynamic pattern formation (Chapter 5 – Figures 5.6D and S5.5A), whereby a moderate amount of noise increases the probability of generating spatially ordered configurations. We should further explore these results by deriving more exact results for realistic noise descriptions, similar to what has been done to study stochastic Turing patterns [Biancalani et al., 2010].

## **2. Relate our model results to experiments (as discussed earlier)**

We have discussed in detail how various experimental approaches can be utilized to assess the conditions under which our results hold. Ideally, we would like our findings to be verified in all three classes of experimental systems proposed above. However, directly testing our ideas in any of these systems poses technical challenges which would need to be overcome. For developmental systems, our aim here would be to simply identify more biological systems exhibiting dynamic phenomena (e.g. traveling waves) that qualitatively match our model findings. For instance, several well-known systems are still modeled using phenomenological models such as the FitzHugh-Nagumo model [Sgro et al., 2015; Hubaud et al., 2017]. Instead, we could construct bottom-up multiscale models from known genetic interactions to try to generate similar phenomena. Secondly, synthetic systems – both living and non-living ones – offer better prospects for running controlled experimental studies that directly test our findings. Arguably the best option here would be to construct a large library of strains with different quorum-sensing modules and screen their population-level dynamics, as mentioned earlier. Finally, non-living synthetic systems could potentially achieve the highest degree of control, but efforts in this direction are currently limited by technical challenges in engineering complex signaling modules.

## **3. Perform network enumeration studies of pattern formation in continuous systems (e.g. reaction-diffusion systems)**

A logical question to ask about the cellular automaton modeling approach we took, is how our results compare to continuous systems, such as reaction-diffusion systems. This is important to address, because we would like to be able to tell apart persistent

features from our data from artefacts arising from our modeling choices. In fact, dynamical systems are known to show different behavior when modeling choices are slightly varied. A well-known example in this context is the logistic map, which shows all hallmarks of chaotic in its discrete formulation, while its continuous counterpart has an unremarkable single stable fixed point [Strogatz, 1994]. In this example, changing time in the model from a continuous variable to a discrete one has drastic consequences to the system's dynamics. For our model, we have seen that changing cells states from discrete (as arising from a step function response) to continuous (as arising from a sigmoidal response function) may also drastically change our results. For instance, at a low enough response function steepness (Hill coefficient), the system always becomes spatially homogeneous, under the same set of parameters that would give dynamic spatial patterns at infinite Hill coefficient (Chapter 5 – Figures S5.5A–B). In a similar vein, we could ask whether going from discrete to continuous time and space have any impact on our findings. However, it is not straightforward how this operation would work in practice, because going from discrete to continuous variables necessarily introduces additional variables (e.g. a response time parameter to characterize how quickly cells respond to signals).

## 6

Another angle to consider is to perform a network enumeration study, similar to ours, to find traveling waves in reaction-diffusion systems. As mentioned, large-scale network screenings for Turing patterns in reaction-diffusion systems do exist, but these focus on static patterns, which can be easily studied by looking for instabilities in the real part of the spectrum [Murray, 2003], which does not even require performing numerical simulations. Potentially such studies could be extended to include oscillations, by looking at the imaginary part of the spectrum. However, it is not clear whether we can screen reaction-diffusion systems for traveling waves in a similar simulation-free manner. Efforts along these lines may potentially benefit from the large body of literature on waves in continuous media [Van Saarloos, 2003].

#### **4. Develop analytic methods that are applicable to a wider range of modeled cases (e.g. gene circuits with repression, multiple genes)**

A significant part of this thesis deals with developing analytic methods to predict the dynamics and steady states of the system, to identify quantitatively different model behaviors or phases and to find conditions for certain behaviors (e.g. traveling wave propagation). While we have succeeded in finding a number of analytic treatments, the current efforts should be expanded to building more general frameworks that are applicable to a wider range of situations.

The “pseudo-energy landscape” is such an analytic method that predicts the system's

dynamics without simulations. However, in its current form it only applies to cells secreting a single molecule with a positive feedback loop. Furthermore, the equation of motion – although it works well – is phenomenological and not rigorously derived from the microscopic rules of the cellular automaton. One direct challenge in extending our current framework to more complex situations is that it is not clear how dynamic patterns should be represented on this “pseudo-energy” landscape. This is true even for oscillations arising from a single negative feedback loop, which would repeatedly cycle between different positions on the energy landscape. However, the self-organization of complex patterns does give a sense of directionality in time (i.e. simulations are clearly different when played backwards). Thus, finding a representation of this “arrow of time” in the self-organization dynamics remains an open challenge (to be discussed further in point 6).

An alternative approach is to directly derive macroscopic equations of motion from our cellular automaton model. This was previously attempted in [Olimpio, 2016], which demonstrated a method to derive equations for the dynamics of the macrostate  $(p, I)$  as a Markovian stochastic process. This relied on a mean-field approach, where one estimates the sensed concentrations of ON and OFF cells as population averages that depend on the degree of spatial order. The results obtain in this way matches well with simulations in certain cases, but showed wild fluctuations and large deviations in other cases, especially in the spatial index  $I$ . As such, we opted for the phenomenological equation of motion derived from the “pseudo-energy” landscape instead, which despite not being rigorously derived, qualitatively appears to do a better job in reproducing macroscopic dynamics.

Similarly, we have attempted to use a similar approach to derive macroscopic equations for a system with two molecules. This yielded a set of self-consistent equations for the fractions of cells in each of the four states  $(0,0)$ ,  $(1,0)$ ,  $(0,1)$  and  $(1,1)$ . These mean-field derived equations qualitatively reproduce the dynamics in limiting cases where the dynamics becomes trivial, such as when all cells acquire the same gene expression state over time. However, it is unable to reproduce any of the more complex phenomena such as long chaotic fluctuations, or formation of dynamic spatial patterns. For this reason, the detailed results of this approach are omitted from this thesis.

Another possibility would be to derive an analytic description of our system as a continuous-space, continuous-time system (e.g. a set of partial differential equations). This would circumvent the need to introduce macroscopic variables (such as  $p$  and  $I$ ) and would allow potentially easier analysis of steady state solutions. For a particular class of cellular automata, namely *lattice gas cellular automata*, there are rigorous methods to derive

partial differential equations from underlying microscopic rules [Deutsch & Dormann, 2005]. This construction relies on a (different type of) mean-field approximation, where one assumes that probability distributions at individual lattice sites are independent. Whether these or other methods can be generalized or adapted to our model is an open question.

### **5. Further develop analytic methods for studying dynamic patterns in discrete systems such as the multicellular model considered in this thesis**

In addition to developing general analytic frameworks for representing the system's dynamics, we should also aim to extend analytic methods that target specific features of our models. For instance, we would ideally like to extend the scope of our traveling wave propagation conditions to different types of dynamic patterns. The conditions only apply to specific waves composed of single bands of cells, but from examples we know that waves can take a variety of morphologies (Chapter 5 – Figure 5.2). So, it would be useful to extend these conditions to include different “wave morphologies”. As a second example, we indirectly infer the presence of waves in simulations from features such as macroscopic variables that do not change over time and a periodicity of the pattern that is a multiple of the grid size (see Chapter 5 – Section S5.4.2). While we have verified that this approach works by testing it on example simulations, it is by no means a complete approach to extract all potentially interesting dynamic spatial patterns in the model.

Therefore, it would be useful to have more analytic methods for analyzing complex dynamic spatial patterns, which are not limited to specific pattern morphologies (e.g. straight traveling waves). Just as linear stability analysis is a powerful standard method to look for Turing instabilities [Murray, 2003], having a universal tool that is employable for a range of discrete models in this context would be very useful. It would allow us to identify interesting patterns, and allow us to screen larger sets of networks, both without examining simulations by eye. Highly formalized results on traveling waves on 2D lattices do exist in the mathematics literature, but have not found their way into biology yet [Hupkes & Van Vleck, 2013]. Furthermore, in optics discrete solitons (waves with fixed shapes traveling at constant velocity) have been studied thoroughly both theoretically and experimentally on photonic lattices [Lederer et al., 2008]. Relating our results to the mathematics and physics literature on waves on 2D lattices may reveal useful methods for studying the dynamic spatial patterns that we observed.

### **6. Characterize and predict trajectories of complex patterns from a dynamical systems perspective**

Another aspect that is insufficiently understood is the self-organization dynamics of complex patterns such as dynamical spatial patterns. Typical studies of pattern formation focus on the steady state patterns, while often neglecting the dynamics of self-organization. This is a natural result of the linear stability analysis done for reaction-diffusion systems, which considers only long-time steady states and neglects dynamics [Murray, 2003]. In our work, we have aimed to extract general characteristics of the self-organization process from batch simulations (Chapter 5 – Section 5.4.2). However, our understanding is currently at the stage of characterization, and we have no tools available to extract more rigorous and potentially predictive results about the self-organization dynamics.

Specifically, we would like to have tools to predict and understand a number of features we observed. First, recall that formation times for traveling waves are roughly exponentially distributed (Chapter 5 – Figure 5.8A). One can then speculate that this may arise from an underlying Markov process – specifically, a Bernoulli process where the chance of forming a wave is equal at every time step. However, whether one can construct a simplified model that reproduces this feature and also gives insight into the mechanism of wave formation is unknown. Second, the long transient chaotic phases preceding formation of a steady state pattern (Chapter 5 – Figure 5.5) are indistinguishable amongst different attractor states. This means that whether the final pattern is a wave or a fixed point without spatial structure, the transient dynamics look qualitatively similar. But could there be hidden determinants, indiscernible by the human eye, that contain information about the long-time dynamics (e.g. whether a wave forms at the end of the simulation or not)? Black-box methods from machine learning could potentially offer answers [Mehta et al., 2019].

Perhaps we should take a broader perspective, and we ask questions about the system from a dynamical systems perspective [Strogatz, 1994]. How many attractors are there in the system? How are these different attractors organized in phase space? What is the stability of each of the attractors? What is the attractor structure of a dynamic spatial pattern? Given the complexity of the transient dynamics before the formation of a steady state dynamic spatial pattern, it is not unreasonable to assume that the attractor to these states could have a highly non-trivial structure. Could it be a strange attractor, with a fractal structure? Furthermore, we know from simulations that the waves have limited stability. While waves can sometimes easily recover from isolated perturbations on single cells, perturbing multiple cells simultaneously often leads to wave breakdown [Grundel, 2019]. However, without tools to analytically determine the steady states or to calculate the stability of attractors it is hard to make any rigorous statements about these points.

## 7. Relate our results to the wider literature on complex systems, computation and thermodynamics

The phenomena we observed in our models show hallmarks of complex systems; non-linearity, feedback, emergence, self-organization and openness. The underlying model is inherently nonlinear through the imposed response function (either step-like or sigmoidal), which creates a feedback loop – every cell senses and responds its environment by adjusting secretion rates. We observe the emergence of non-trivial patterns, which are self-organized from random initial conditions. Furthermore, we note that our system is open in the sense that it requires a flow of energy from the environment to be sustained. While this is not explicitly modeled, cells require a constant energy source to stay alive and express genes. Altogether, this implies that we should aim to place our work in the wider context of studies on complex systems in biology [Kaneko, 2006].

On the other hand, there are also traits which are often associated with complex systems that are not present in our system. First, the network structure of our model is trivial: all cells are in principle coupled to each other, making our system a fully connected network. However, as mentioned before, in real tissues systems of interacting cells may form complex networks with nontrivial topologies [Gosak et al., 2018]. Second, our system has limited adaptability, because it has no real memory. The cells are adaptive in the sense that they have the capacity to respond to their environment. However, cellular automata are by definition Markovian – the next state determines only on the present state and not on its history. Expanding our model to consider how adjusting these two additional features – network structure and adaptability, which requires memory – is another logical extension of our work.

Another interesting perspective to consider is whether we can treat a collection of interacting cells as an abstract computational machine. Cellular automata are known to be capable of performing computations by simulating abstract computational machine. It has been suggested that this ability arises at the edge between an ordered phase and a chaotic phase, although these claims remain controversial [Langton, 1990; Mitchell et al., 1993]. Furthermore, the classic Game of Life has been shown to be a universal Turing machine, capable of universal computation [Berlekamp et al., 2004; Rendell, 2011]. Likewise, we may ask whether any computational tasks can be performed using our setup of interacting cells. Since we can produce traveling waves and other dynamic spatial patterns, we may wonder whether there is any way of combining these patterns to perform basic logic operations, by setting up colliding patterns that interact in required manners.

Finally, it would be interesting to relate our work to non-equilibrium statistical physics and thermodynamics. Several decades ago, researchers showed that systems far from

thermodynamic equilibrium can show remarkable examples of self-organization, such as hexagonal patterns when a fluid is heated beyond an instability threshold in convection cells (known as Bénard cells) [Prigogine & Stengers, 1984]. These systems, known as dissipative structures, follow thermodynamic principles different from equilibrium systems, and even systems close to equilibrium, for which general principles such as minimum entropy production and Onsager's reciprocal relations can still be obtained [Prigogine, 1977]. These findings are highly relevant to living systems, as they are inherently out of equilibrium [Phillips, 2015]. Also, both energy dissipation and self-organization are ubiquitous in biology. Recent work has proposed novel thermodynamic frameworks for processes in living systems such as self-replication [England, 2013], adaptation [Perunov et al., 2016] and self-assembly [England, 2015].

Relating our system to thermodynamics would allow us to address questions such as whether there are physical principles constraining self-organization of spatial patterns, or even development in general? Furthermore, are there extremal principles similar to minimum entropy production [Lebon et al., 2008], that can explain the self-organization phenomena observed in our system? While our “pseudo-energy landscape” is an exploration that is inspired by such approaches, it is not a thermodynamic construction and does not rely on any exact correspondence to any thermodynamic system. One way to proceed, therefore, is to try to find such an exact correspondence. For instance, it has been established that reaction-diffusion systems, when modeled as probabilistic cellular automata, can in certain cases be mapped onto Ising spin models [Weimar and Boon, 1994; Merle et al., 2019]. These studies show that it is possible to make surprising new connections between physics and biology. Ultimately, searching for such unexpected links and hidden connections will bring us closer to a physics-type understanding of living systems.

# ACKNOWLEDGEMENTS

Research is never a truly solitary activity, even if most of the work is done independently. As a theorist in a mostly experimental research group, I still benefited from numerous interactions with colleagues, friends and family, both within the department and outside. Although only a small number of people were directly involved in shaping the work in this thesis, the remaining people to acknowledge have all in some way or another influenced me and therefore my work in the past few years.

First and foremost, I would like to thank my supervisor Hyun Youk. From the beginning, you were a very dedicated supervisor and mentor, guiding me through my transition from theoretical physics into biology. The initial phase of my PhD was far from smooth, but you have always managed to guide me and my PhD forward, without directly pulling the ropes for me. Also thanks for your flexibility in allowing me to work on my own project of interest. You have also really changed the way I view research over the course of these years. As an incoming PhD student with a theoretical physics background, I initially thought research was all about doing esoteric work, incomprehensible to ordinary people, by digging deeper and deeper in a highly specialized field. However, I now see the value of starting with an easy-to-understand phenomenon and to explore it further and further, until we eventually gain unexpected new insights. Your mentorship and involvement in my career progression have also been indispensable, so that I am now ready to take the next step of my career and life.

I would like to thank all members of the committee for taking time to read my thesis. My apologies for the length of this book. If I ever get to supervise a PhD student, I will advise him/her to write a more concise thesis than mine.

Next, I would like to acknowledge many other principal investigators (PIs) who have supported me and provided feedback. To begin with, I would like to thank my promotor Marileen Dogterom, also as head of the department, for providing a good working environment for me to conduct my research in. Next, many thanks to Martin Depken, one of the two theory PIs at BN, for giving feedback on my work and progress. It was great discussing science, careers and life as a BN-theorist with you. Across the campus, at the Quantum Nanoscience department, I also benefitted from discussing with Yaroslav Blanter. Thank you for taking the time to learn about my work and asking interesting

questions. During the four years, I was also involved in teaching courses of the Nanobiology program. Thanks to Chirlmin Joo and Liedewij Laan for giving me the opportunity to not only correct endless exams and problem sets, but also teach a few lectures, interact with students and design exercises. Finally, thanks to Marie-Eve, Timon, Greg, Cees, Dimphna and Bertus for various nice and useful interactions we had over the course of the years.

A very important group of people to thank are my lab mates. Thanks for sitting through an endless number of my group meeting and other presentations, and for providing critical feedback. Although you worked on very different projects, it was also great to be exposed to many different topics and engage with such a diverse group. More importantly, thanks for providing the social environment that defined our lab's culture during all these years! Whether it was cracking lame jokes in the office, getting stranded at the Paris airport or coming up with the Youksters, we surely had many fun moments in the midst of conducting serious scientific research.

Let me start by thanking three of you who were in one way or another involved in the work in this thesis. First, I acknowledge the pioneering work by Théo Maire, who established the model that became the basis of my thesis. Although you became a spore guy after re-entering as a PhD student, your continued interest has spurred discussions that have improved the quality of my work. Second, thanks to Eduardo Olimpio for his brilliance in establishing a formal mathematical framework of Théo's model. It was impressive what you achieved in a master project of a few months, and continuing your work was the main driving force behind the latter part of my PhD. Finally, many thanks to the only student I supervised, Douwe Grundel. Thanks for your input and hard work during your master internship! In retrospect, it appears that my most productive period were the months you were in the lab. Having someone to discuss with every day definitely gave our project a huge boost in creativity.

Naturally, I also thank everyone else who has been part of the lab at some point of my PhD. Mehran, we were the brave first PhD students to join the lab and made it through till the end, so congrats to ourselves! Diego, thanks for teaching me basic lab skills when I joined the lab. Also, it was fun speaking multiple languages together and sorry I did not continue with salsa dancing. Hiram, it was never boring around you; thanks for bringing life and humor into the group. Lars, great to have a real biologist around and as the only dad in our group you created a family(-friendly) environment. Diederik, thanks for your involvements in group meetings and for being the other maths nerd in the lab. Milan, it was great to do activities together such as skiing in Courchevel and thanks for introducing me to bouldering. Also thanks to Eve for the fun and Yuliia for the nice conversations,

and I hope you both are doing well. Finally, it was a pleasure to share a lab with the many students and technicians that passed the lab at some point, including Max, Jasper, Ilya, Berkalp, Arie, Tim, Pim, Raymond, Laura, Stefan, David, Marloes and Jacinta.

My gratitude extends to the rest of the BN department for the collegiality and generally relaxed atmosphere. It would be risky to try to list all of you individually, as I am bound to forget people. Let me therefore first thank all of you, and now give a special mention to a few who have made a lasting impression on me. The first special mention goes to all the wonderful BN musicians I have played with at numerous occasions, from Quo Vadis to the Van der Does concert in the atrium of the building. Thanks Margreet, Christine, Fabai, Helena and many others that were involved in one of the concerts. Margreet, bassoon and piano make a great combination and it is a pity we will not be able to continue playing. Next, I want to thank to my fellow BN theorists Misha, Afshin and George for making me feel like a less rare breed of animal in the department. Special thanks to Misha, for being helpful for anything from discussing thesis submission to holding sheet music during a windy Quo Vadis concert. Another special thanks to everyone from the Laan lab and the Bokinsky lab. We shared a wet lab with you guys (in the old and new buildings respectively), which naturally meant we also shared many experiences and social activities — this is even true for a theorist like myself. Finally, a big applause to the secretaries, especially our ‘own’ secretaries Chantal, Dijana and Nadine, for always being helpful and responsive.

Although as a PhD student it is easy to live inside a TU Delft bubble, or worse, a BN department bubble, I still managed to have contact with the outside world. Sharing an apartment was a natural way to break the bubble. In particular, I would like thank to my (frequently changing) housemates at the Hippolytusbuurt for the nice dinners (usually risotto) and random encounters in the house.

Another group to thank are my university friends from Utrecht: Thomas, Remco, Tim, Henri and Stefan. Although we have had less contact since graduating, we have managed to meet on a regular basis. Great to keep up the reunions and thanks being interested in what I am doing now, even more so than some of my lab mates.

My final word of thanks goes to my parents. You have always supported and encouraged me unconditionally. I am lucky to have such great parents living nearby. 爸爸妈妈，谢谢你们的支持和鼓励！

# List of Figures

1.1	Biological examples of self-organization . . . . .	3
2.1	Different modes of chemical signaling . . . . .	12
2.2	Non-chemical pattern formation . . . . .	15
2.3	Mechanisms of pattern formation . . . . .	22
3.1	Schematic of our agent-based multicellular model in its original formulation . . . . .	33
3.2	Finite Hill coefficient: example simulations showing the transition from collectiveness to autonomy . . . . .	40
3.3	Finite Hill coefficient: Global phase diagram, for the specific case that the system is uniform . . . . .	44
3.4	Finite Hill coefficient: Phase diagrams showing steady states with their stability . . . . .	45
3.5	Finite Hill coefficient: Transition between the collective phase and the autonomous phase . . . . .	47
3.6	Finite Hill coefficient: Dynamics of metrics to characterize the autonomy-collectiveness transition . . . . .	52
3.7	Finite Hill coefficient: Mean-field analysis of steady states . . . . .	56
3.8	Finite Hill coefficient: Stability of the uniform lattice solution for the bistable system with $n = 2$ . . . . .	61
3.9	Sensitivity to initial conditions: Perturbation response of a population of digital cells . . . . .	65
3.10	Sensitivity to initial conditions: Perturbation response of a population of analogue cells . . . . .	66
3.11	Mutual information (binary cells): relation to entropy of population . . . . .	75
3.12	Mutual information (binary cells): dependence on lattice spacing . . . . .	76
3.13	Mutual information (binary cells): effect of adding noise . . . . .	79
3.14	Mutual information (analogue cells): effect of adding noise . . . . .	82

4.1	Statistical dynamics: Secrete-and-sense cells can be classified into distinct behavioral phases . . . . .	89
4.2	Statistical dynamics: Spatial configurations of secrete-and-sense cells (microstates) can be grouped into macrostates. . . . .	91
4.3	Statistical dynamics: A cellular lattice acts as a particle that rolls down on and adheres to a pseudo-energy landscape . . . . .	93
4.4	Statistical dynamics: Gradient field of the pseudo-energy and the trapping probability completely specify the particle's trajectory. . . . .	95
4.5	Statistical dynamics: Stochastic sensing can yield spatial configurations that are more ordered than those formed without noise . . . . .	97
S4.1	Density of states as a function of the macroscopic variables ( $p, I$ ) . . . . .	103
S4.2	Gradient fields of the pseudo-energy and particle trajectories, in the case of weak interactions . . . . .	104
S4.3	Gradient fields of the pseudo-energy and particle trajectories, in the case of strong interactions . . . . .	105
5.1	Computationally screening cellular dialogues to find ones that enable dynamic patterns to form . . . . .	129
5.2	Examples of self-organized dynamic patterns found through computational screening . . . . .	131
5.3	Computational search revealed tree structures that group cellular dialogues based on their ability to generate either static patterns, dynamic temporal patterns, or dynamic spatial patterns . . . . .	135
5.4	Analytic framework predicts and explains how cells can sustain dynamic spatial patterns . . . . .	138
5.5	Three-step, "order-fluctuate-settle" process leads to formation of dynamic spatial patterns . . . . .	142
5.6	Dynamic spatial patterns in an extended model with more complex elements of cell-cell communication . . . . .	146
5.7	Self-organized dynamic-pattern-forming systems with poorly understood interactions that our theoretical framework may help in elucidating . . . . .	149
S5.1	Statistics of trajectories from each of the three classes of cellular dialogues	151
S5.2	Self-organization of static and dynamic patterns for cellular dialogues with OR-logic signal integration, instead of AND-logic . . . . .	153
S5.3	Additional examples showing how the two macroscopic variables vary over time during self-organization of dynamic spatial patterns . . . . .	159
S5.4	Algorithm used to generate spatial configurations with a desired value of spatial index and a fraction of cells which are ON. . . . .	161

---

S5.5	Formation and propagation of traveling waves under the influence of noisy gene-expression, finite Hill coefficient, and spatially changing parameter values . . . . .	166
S5.6	Dynamic patterns with oscillating cells and their associated state diagrams	175
S5.7	State diagrams — a directed graph representation of dynamics at the single-cell level . . . . .	182
S5.8	Statistics on how long cells take to form traveling waves . . . . .	207
S5.9	Comparison between simulations and analytic theory for traveling wave propagation . . . . .	209
S5.10	Robustness of traveling waves, measured by the fraction of parameter sets capable of propagating traveling waves . . . . .	212
S5.11	Analytic framework predicts robustness of wave formations . . . . .	220
S5.12	Reliability of traveling wave formation . . . . .	222
S5.13	Persistence of traveling waves in the extended model with stochasticity: theory and simulations . . . . .	226

# REFERENCES

- <sup>1</sup>M. Ackermann, “A functional perspective on phenotypic heterogeneity in microorganisms”, *Nature Reviews Microbiology* **13**, 497 (2015).
- <sup>2</sup>L. Anselin, *Spatial Econometrics: Methods and Models*, Studies in Operational Regional Science (Springer, Dordrecht, 1988).
- <sup>3</sup>Y. E. Antebi, S. Reich-Zeliger, Y. Hart, A. Mayo, I. Eizenberg, J. Rimer, P. Putheti, D. Pe’er, and N. Friedman, “Mapping Differentiation under Mixed Culture Conditions Reveals a Tunable Continuum of T-Cell Fates”, *PLOS Biology* **11**, e1001616 (2013).
- <sup>4</sup>B. Alberts, A. D. Johnson, J. Lewis, D. Morgan, M. Raff, K. Roberts, and P. Walter, *Molecular Biology of the Cell*, 5th, Molecular Biology of the Cell (Garland Science, 2007).
- <sup>5</sup>U. Alon, *An Introduction to Systems Biology: Design Principles of Biological Circuits*, 1st ed. (CRC Press, Boca Raton, FL, 2006).
- <sup>6</sup>S. Alt, P. Ganguly, and G. Salbreux, “Vertex models: from cell mechanics to tissue morphogenesis”, *Philosophical Transactions of the Royal Society B: Biological Sciences* **372**, 20150520 (2017).
- <sup>7</sup>M. Amos, “Population-based microbial computing: a third wave of synthetic biology?”, *International Journal of General Systems* **43**, 770–782 (2014).
- <sup>8</sup>P. W. Anderson, “More Is Different”, *Science* **177**, 393 (1972).
- <sup>9</sup>K. Aoki, Y. Kondo, H. Naoki, T. Hiratsuka, R. E. Itoh, and M. Matsuda, “Propagating Wave of ERK Activation Orients Collective Cell Migration”, *Developmental Cell* **43**, 305–317.e5 (2017).
- <sup>10</sup>S. Artavanis-Tsakonas, M. D. Rand, and R. J. Lake, “Notch Signaling: Cell Fate Control and Signal Integration in Development”, *Science* **284**, 770 (1999).
- <sup>11</sup>A. K. Baidya, S. Bhattacharya, G. P. Dubey, G. Mamou, and S. Ben-Yehuda, “Bacterial nanotubes: a conduit for intercellular molecular trade”, *Current Opinion in Microbiology* **42**, 1–6 (2018).
- <sup>12</sup>R. E. Baker, S. Schnell, and P. K. Maini, “Mathematical Models for Somite Formation”, in *Current Topics in Developmental Biology*, Vol. 81, edited by S. Schnell, P. K. Maini, S. A. Newman, and T. J. Newman (Academic Press, 2008), pp. 183–203.
- <sup>13</sup>K. Barros, R. Chacko, J. Gould, H. Gould, N. Gulbahce, J. Sibley Peter Tobochnik, and H. Wang, *Density of States of the 2D Ising Model*, (2015) <http://stp.clarku.edu/simulations/ising/wanglandau/index.html>.

- <sup>14</sup>Y. Bar-Yam, *Dynamics Of Complex Systems*, 1st ed. (Avalon Publishing, 2003).
- <sup>15</sup>M. Belting and A. Wittrup, “Nanotubes, exosomes, and nucleic acid-binding peptides provide novel mechanisms of intercellular communication in eukaryotic cells: implications in health and disease”, *The Journal of Cell Biology* **183**, 1187 (2008).
- <sup>16</sup>H. C. Berg and E. M. Purcell, “Physics of chemoreception”, *Biophysical Journal* **20**, 193–219 (1977).
- <sup>17</sup>E. R. Berlekamp, J. H. Conway, and R. K. Guy, *Winning Ways for your Mathematical Plays*, 2nd ed., Vol. 4 (A. K. Peters Ltd., Wellesley, Massachusetts, 2001–2004).
- <sup>18</sup>T. Biancalani, D. Fanelli, and F. Di Patti, “Stochastic Turing patterns in the Brusselator model”, *Physical Review E* **81**, 046215 (2010).
- <sup>19</sup>J. S. Bois, F. Jülicher, and S. W. Grill, “Pattern Formation in Active Fluids”, *Physical Review Letters* **106**, 028103 (2011).
- <sup>20</sup>B. Bolker and S. W. Pacala, “Using Moment Equations to Understand Stochastically Driven Spatial Pattern Formation in Ecological Systems”, *Theoretical Population Biology* **52**, 179–197 (1997).
- <sup>21</sup>F. Brinkmann, M. Mercker, T. Richter, and A. Marciniak-Czochra, “Post-Turing tissue pattern formation: Advent of mechanochemistry”, *PLOS Computational Biology* **14**, e1006259 (2018).
- <sup>22</sup>J. Briscoe and A. Kicheva, “The physics of development 100 years after D’Arcy Thompson’s “On Growth and Form””, *Mechanisms of Development* **145**, 26–31 (2017).
- <sup>23</sup>N. E. Buchler, U. Gerland, and T. Hwa, “On schemes of combinatorial transcription logic”, *Proceedings of the National Academy of Sciences* **100**, 5136 (2003).
- <sup>24</sup>E. O. Budrene and H. C. Berg, “Complex patterns formed by motile cells of *Escherichia coli*”, *Nature* **349**, 630–633 (1991).
- <sup>25</sup>D. E. Cameron, C. J. Bashor, and J. J. Collins, “A brief history of synthetic biology”, *Nature Reviews Microbiology* **12**, 381 (2014).
- <sup>26</sup>V. Castets, E. Dulos, J. Boissonade, and P. De Kepper, “Experimental evidence of a sustained standing Turing-type nonequilibrium chemical pattern”, *Physical Review Letters* **64**, 2953–2956 (1990).
- <sup>27</sup>J. B. Chang and J. E. Ferrell Jr, “Mitotic trigger waves and the spatial coordination of the *Xenopus* cell cycle”, *Nature* **500**, 603 (2013).
- <sup>28</sup>G. Chatterjee, N. Dalchau, R. A. Muscat, A. Phillips, and G. Seelig, “A spatially localized architecture for fast and modular DNA computing”, *Nature Nanotechnology* **12**, 920 (2017).
- <sup>29</sup>A. H. Chau, J. M. Walter, J. Gerardin, C. Tang, and W. A. Lim, “Designing Synthetic Regulatory Networks Capable of Self-Organizing Cell Polarization”, *Cell* **151**, 320–332 (2012).

- <sup>30</sup>C.-C. Chen, L. Wang, M. V. Plikus, T. X. Jiang, P. J. Murray, R. Ramos, C. F. Guerrero-Juarez, M. W. Hughes, O. K. Lee, S. Shi, R. B. Widelitz, A. D. Lander, and C. M. Chuong, “Organ-Level Quorum Sensing Directs Regeneration in Hair Stem Cell Populations”, *Cell* **161**, 277–290 (2015).
- <sup>31</sup>X. Cheng and J. E. Ferrell, “Apoptosis propagates through the cytoplasm as trigger waves”, *Science* **361**, 607 (2018).
- <sup>32</sup>J. Cooke and E. C. Zeeman, “A clock and wavefront model for control of the number of repeated structures during animal morphogenesis”, *Journal of Theoretical Biology* **58**, 455–476 (1976).
- <sup>33</sup>M. Coppey, A. M. Berezhevskii, S. C. Sealfon, and S. Y. Shvartsman, “Time and Length Scales of Autocrine Signals in Three Dimensions”, *Biophysical Journal* **93**, 1917–1922 (2007).
- <sup>34</sup>F. Corson, L. Couturier, H. Rouault, K. Mazouni, and F. Schweisguth, “Self-organized Notch dynamics generate stereotyped sensory organ patterns in *Drosophila*”, *Science* **356**, eaai7407 (2017).
- <sup>35</sup>J. Cotterell, A. Robert-Moreno, and J. Sharpe, “A Local, Self-Organizing Reaction-Diffusion Model Can Explain Somite Patterning in Embryos”, *Cell Systems* **1**, 257–269 (2015).
- <sup>36</sup>J. Cotterell and J. Sharpe, “An atlas of gene regulatory networks reveals multiple three-gene mechanisms for interpreting morphogen gradients”, *Molecular Systems Biology* **6**, 425 (2010).
- <sup>37</sup>T. Danino, O. Mondragón-Palomino, L. Tsimring, and J. Hasty, “A synchronized quorum of genetic clocks”, *Nature* **463**, 326 (2010).
- <sup>38</sup>J. Davies, “Using synthetic biology to explore principles of development”, *Development* **144**, 1146 (2017).
- <sup>39</sup>K. Deisseroth, “Optogenetics”, *Nature Methods* **8**, 26 (2010).
- <sup>40</sup>V. E. Deneke and S. Di Talia, “Chemical waves in cell and developmental biology”, *The Journal of Cell Biology* **217**, 1193 (2018).
- <sup>41</sup>A. Deutsch and S. Dormann, *Cellular Automaton Modeling of Biological Pattern Formation*, 1st ed., Modeling and Simulation in Science, Engineering and Technology (Birkhäuser Basel, Basel, Switzerland, 2005).
- <sup>42</sup>X. Diego, L. Marcon, P. Müller, and J. Sharpe, “Key Features of Turing Systems are Determined Purely by Network Topology”, *Physical Review X* **8**, 021071 (2018).
- <sup>43</sup>D. E. Discher, P. Janmey, and Y.-I. Wang, “Tissue Cells Feel and Respond to the Stiffness of Their Substrate”, *Science* **310**, 1139 (2005).
- <sup>44</sup>B. A. Doğaner, L. K. Q. Yan, and H. Youk, “Autocrine Signaling and Quorum Sensing: Extreme Ends of a Common Spectrum”, *Trends in Cell Biology* **26**, 262–271 (2016).
- <sup>45</sup>W. Driever and C. Nüsslein-Volhard, “The bicoid protein is a positive regulator of hunchback transcription in the early *Drosophila* embryo”, *Nature* **337**, 138–143 (1989).

- <sup>46</sup>G. P. Dubey and S. Ben-Yehuda, “Intercellular Nanotubes Mediate Bacterial Communication”, *Cell* **144**, 590–600 (2011).
- <sup>47</sup>E. Dubuc, P. A. Pieters, A. J. van der Linden, J. C. M. van Hest, W. T. S. Huck, and T. F. A. de Greef, “Cell-free microcompartmentalised transcription–translation for the prototyping of synthetic communication networks”, *Current Opinion in Biotechnology* **58**, 72–80 (2019).
- <sup>48</sup>G. A. Dunn and A. F. Brown, “A Unified Approach to Analysing Cell Motility”, *Journal of Cell Science* **1987**, 81 (1987).
- <sup>49</sup>J. L. England, “Statistical physics of self-replication”, *The Journal of Chemical Physics* **139**, 121923 (2013).
- <sup>50</sup>J. L. England, “Dissipative adaptation in driven self-assembly”, *Nature Nanotechnology* **10**, 919 (2015).
- <sup>51</sup>E. V. Entchev, A. Schwabedissen, and M. González-Gaitán, “Gradient Formation of the TGF- $\beta$  Homolog Dpp”, *Cell* **103**, 981–992 (2000).
- <sup>52</sup>T. Erdmann, M. Howard, and P. R. ten Wolde, “Role of Spatial Averaging in the Precision of Gene Expression Patterns”, *Physical Review Letters* **103**, 258101 (2009).
- <sup>53</sup>G. B. Ermentrout and L. Edelstein-Keshet, “Cellular Automata Approaches to Biological Modeling”, *Journal of Theoretical Biology* **160**, 97–133 (1993).
- <sup>54</sup>J. Eyckmans, T. Boudou, X. Yu, and C. S. Chen, “A Hitchhiker’s Guide to Mechanobiology”, *Developmental Cell* **21**, 35–47 (2011).
- <sup>55</sup>S. Fancher and A. Mugler, “Fundamental Limits to Collective Concentration Sensing in Cell Populations”, *Physical Review Letters* **118**, 078101 (2017).
- <sup>56</sup>J. E. Ferrell and S. H. Ha, “Ultrasensitivity part I: Michaelian responses and zero-order ultrasensitivity”, *Trends in Biochemical Sciences* **39**, 496–503 (2014a).
- <sup>57</sup>J. E. Ferrell, Jr, and S. H. Ha, “Ultrasensitivity part II: multisite phosphorylation, stoichiometric inhibitors, and positive feedback”, *Trends in Biochemical Sciences* **39**, 556–569 (2014b).
- <sup>58</sup>J. E. Ferrell and S. H. Ha, “Ultrasensitivity part III: cascades, bistable switches, and oscillators”, *Trends in Biochemical Sciences* **39**, 612–618 (2014c).
- <sup>59</sup>R. FitzHugh, “Impulses and Physiological States in Theoretical Models of Nerve Membrane”, *Biophysical Journal* **1**, 445–466 (1961).
- <sup>60</sup>A. G. Fletcher, M. Osterfield, R. E. Baker, and S. Y. Shvartsman, “Vertex Models of Epithelial Morphogenesis”, *Biophysical Journal* **106**, 2291–2304 (2014).
- <sup>61</sup>P. Fried, M. Sánchez-Aragón, D. Aguilar-Hidalgo, B. Lehtinen, F. Casares, and D. Iber, “A Model of the Spatio-temporal Dynamics of *Drosophila* Eye Disc Development”, *PLOS Computational Biology* **12**, e1005052 (2016).

- <sup>62</sup>N. Friedman, L. Cai, and X. S. Xie, “Linking Stochastic Dynamics to Population Distribution: An Analytical Framework of Gene Expression”, *Physical Review Letters* **97**, 168302 (2006).
- <sup>63</sup>J. Garcia-Ojalvo, M. B. Elowitz, and S. H. Strogatz, “Modeling a synthetic multicellular clock: Repressilators coupled by quorum sensing”, *Proceedings of the National Academy of Sciences of the United States of America* **101**, 10955 (2004).
- <sup>64</sup>J. Garcia-Ojalvo, “Physical approaches to the dynamics of genetic circuits: a tutorial”, *Contemporary Physics* **52**, 439–464 (2011).
- <sup>65</sup>L. Gelens, G. A. Anderson, and J. E. Ferrell, “Spatial trigger waves: positive feedback gets you a long way”, *Molecular Biology of the Cell* **25**, 3486–3493 (2014).
- <sup>66</sup>A. Getis and J. K. Ord, “The Analysis of Spatial Association by Use of Distance Statistics”, *Geographical Analysis* **24**, 189–206 (1992).
- <sup>67</sup>N. Geva-Zatorsky, N. Rosenfeld, S. Itzkovitz, R. Milo, A. Sigal, E. Dekel, T. Yarnitzky, Y. Liron, P. Polak, G. Lahav, and U. Alon, “Oscillations and variability in the p53 system”, *Molecular Systems Biology* **2**, 2006.0033 (2006).
- <sup>68</sup>F. Ginelli, “The Physics of the Vicsek model”, *The European Physical Journal Special Topics* **225**, 2099–2117 (2016).
- <sup>69</sup>I. Golding and E. C. Cox, “Physical Nature of Bacterial Cytoplasm”, *Physical Review Letters* **96**, 098102 (2006).
- <sup>70</sup>D. T. Gonzales, T.-Y. D. Tang, and C. Zechner, “Moment-based analysis of biochemical networks in a heterogeneous population of communicating cells”, (2019).
- <sup>71</sup>M. Gosak, R. Markovič, J. Dolenšek, M. Slak Rupnik, M. Marhl, A. Stožer, and M. Perc, “Network science of biological systems at different scales: A review”, *Physics of Life Reviews* **24**, 118–135 (2018).
- <sup>72</sup>P. D. Gould, M. Domijan, M. Greenwood, I. T. Tokuda, H. Rees, L. Kozma-Bognar, A. J. W. Hall, and J. C. W. Locke, “Coordination of robust single cell rhythms in the *Arabidopsis* circadian clock via spatial waves of gene expression”, *eLife* **7**, e31700 (2018).
- <sup>73</sup>K. Gousset, E. Schiff, C. Langevin, Z. Marijanovic, A. Caputo, D. T. Browman, N. Chenouard, F. de Chaumont, A. Martino, J. Enninga, J.-C. Olivo-Marin, D. Männel, and C. Zurzolo, “Prions hijack tunnelling nanotubes for intercellular spread”, *Nature Cell Biology* **11**, 328 (2009).
- <sup>74</sup>F. Graner and J. A. Glazier, “Simulation of biological cell sorting using a two-dimensional extended Potts model”, *Physical Review Letters* **69**, 2031–2034 (1992).
- <sup>75</sup>J. Green, “Morphogen gradients, positional information, and *Xenopus*: Interplay of theory and experiment”, *Developmental Dynamics* **225**, 392–408 (2002).
- <sup>76</sup>T. Gregor, W. Bialek, R. R. d. R. van Steveninck, D. W. Tank, and E. F. Wieschaus, “Diffusion and scaling during early embryonic pattern formation”, *Proceedings of the National Academy of Sciences of the United States of America* **102**, 18403 (2005).

- <sup>77</sup>T. Gregor, K. Fujimoto, N. Masaki, and S. Sawai, “The Onset of Collective Behavior in Social Amoebae”, *Science* **328**, 1021 (2010).
- <sup>78</sup>T. Gregor, E. F. Wieschaus, A. P. McGregor, W. Bialek, and D. W. Tank, “Stability and Nuclear Dynamics of the Bicoid Morphogen Gradient”, *Cell* **130**, 141–152 (2007a).
- <sup>79</sup>T. Gregor, D. W. Tank, E. F. Wieschaus, and W. Bialek, “Probing the Limits to Positional Information”, *Cell* **130**, 153–164 (2007b).
- <sup>80</sup>R. K. Grosberg and R. R. Strathmann, “The Evolution of Multicellularity: A Major Transition?”, *Annual Review of Ecology, Evolution, and Systematics* **38**, 621–654 (2007).
- <sup>81</sup>D. A. Grundel, “Investigating travelling waves in a system of communicating cells”, MSc thesis of minor internship (Delft University of Technology, 2019).
- <sup>82</sup>G. Guglielmi, H. J. Falk, and S. De Renzis, “Optogenetic Control of Protein Function: From Intracellular Processes to Tissue Morphogenesis”, *Trends in Cell Biology* **26**, 864–874 (2016).
- <sup>83</sup>G. Haimovich, C. M. Ecker, M. C. Dunagin, E. Eggan, A. Raj, J. E. Gerst, and R. H. Singer, “Intercellular mRNA trafficking via membrane nanotube-like extensions in mammalian cells”, *Proceedings of the National Academy of Sciences* **114**, E9873 (2017).
- <sup>84</sup>H. Haken, *Synergetics: Introduction and Advanced Topics* (Springer-Verlag Berlin Heidelberg, 2004).
- <sup>85</sup>V. Hakim and P. Silberzan, “Collective cell migration: a physics perspective”, *Reports on Progress in Physics* **80**, 076601 (2017).
- <sup>86</sup>J. Halatek and E. Frey, “Rethinking pattern formation in reaction-diffusion systems”, *Nature Physics* **14**, 507–514 (2018).
- <sup>87</sup>E. Hannezo and C.-P. Heisenberg, “Mechanochemical Feedback Loops in Development and Disease”, *Cell* **178**, 12–25 (2019).
- <sup>88</sup>E. Hannezo, C. L. G. J. Scheele, M. Moad, N. Drogo, R. Heer, R. V. Sampogna, J. van Rheenen, and B. D. Simons, “A Unifying Theory of Branching Morphogenesis”, *Cell* **171**, 242–255.e27 (2017).
- <sup>89</sup>E. Hannezo and B. D. Simons, “Statistical theory of branching morphogenesis”, *Development, Growth & Differentiation* **60**, 512–521 (2018).
- <sup>90</sup>Y. Hart, S. Reich-Zeliger, Y. E. Antebi, I. Zaretsky, A. E. Mayo, U. Alon, and N. Friedman, “Paradoxical Signaling by a Secreted Molecule Leads to Homeostasis of Cell Levels”, *Cell* **158**, 1022–1032 (2014).
- <sup>91</sup>I. Heemskerk, K. Burt, M. Miller, S. Chhabra, M. C. Guerra, L. Liu, and A. Warmflash, “Rapid changes in morphogen concentration control self-organized patterning in human embryonic stem cells”, *eLife* **8**, e40526 (2019).
- <sup>92</sup>A. V. Hill, “The possible effects of the aggregation of the molecules of haemoglobin on its oxygen dissociation curve”, *The Journal of Physiology* **40**, iv–vii (1910).

- <sup>93</sup>T. W. Hiscock and S. G. Megason, “Orientation of Turing-like Patterns by Morphogen Gradients and Tissue Anisotropies”, *Cell Systems* **1**, 408–416 (2015).
- <sup>94</sup>P. Hogeweg, “Evolving Mechanisms of Morphogenesis: on the Interplay between Differential Adhesion and Cell Differentiation”, *Journal of Theoretical Biology* **203**, 317–333 (2000).
- <sup>95</sup>J. J. Hopfield, “Neural networks and physical systems with emergent collective computational abilities”, *Proceedings of the National Academy of Sciences* **79**, 2554 (1982).
- <sup>96</sup>J. Howard, S. W. Grill, and J. S. Bois, “Turing’s next steps: the mechanochemical basis of morphogenesis”, *Nature Reviews Molecular Cell Biology* **12**, 392 (2011).
- <sup>97</sup>C. Y. Huang and J. E. Ferrell, “Ultrasensitivity in the mitogen-activated protein kinase cascade”, *Proceedings of the National Academy of Sciences* **93**, 10078 (1996).
- <sup>98</sup>A. Hubaud and O. Pourquié, “Signalling dynamics in vertebrate segmentation”, *Nature Reviews Molecular Cell Biology* **15**, 709 (2014).
- <sup>99</sup>A. Hubaud, I. Regev, L. Mahadevan, and O. Pourquié, “Excitable Dynamics and Yap-Dependent Mechanical Cues Drive the Segmentation Clock”, *Cell* **171**, 668–682.e11 (2017).
- <sup>100</sup>H. J. Hupkes and E. S. Van Vleck, “Negative Diffusion and Traveling Waves in High Dimensional Lattice Systems”, *SIAM Journal on Mathematical Analysis* **45**, 1068–1135 (2013).
- <sup>101</sup>J. Hurtig, D. T. Chiu, and B. Önfelt, “Intercellular nanotubes: insights from imaging studies and beyond”, *Wiley Interdisciplinary Reviews: Nanomedicine and Nanobiotechnology* **2**, 260–276 (2010).
- <sup>102</sup>T. Idema, J. O. Dubuis, L. Kang, M. L. Manning, P. C. Nelson, T. C. Lubensky, and A. J. Liu, “The Syncytial *Drosophila* Embryo as a Mechanically Excitable Medium”, *PLOS ONE* **8**, e77216 (2013).
- <sup>103</sup>T. Idema and A. J. Liu, “Mechanical signaling via nonlinear wavefront propagation in a mechanically excitable medium”, *Physical Review E* **89**, 062709 (2014).
- <sup>104</sup>A. Joesaar, S. Yang, B. Bögels, A. van der Linden, P. Pieters, B. V. V. S. P. Kumar, N. Dalchau, A. Phillips, S. Mann, and T. F. A. de Greef, “DNA-based communication in populations of synthetic protocells”, *Nature Nanotechnology* **14**, 369–378 (2019).
- <sup>105</sup>D. J. Jörg, E. E. Caygill, A. E. Hakes, E. G. Contreras, A. H. Brand, and B. D. Simons, “The proneural wave in the *Drosophila* optic lobe is driven by an excitable reaction-diffusion mechanism”, *eLife* **8**, e40919 (2019).
- <sup>106</sup>K. Kamino, Y. Kondo, A. Nakajima, M. Honda-Kitahara, K. Kaneko, and S. Sawai, “Fold-change detection and scale invariance of cell–cell signaling in social amoeba”, *Proceedings of the National Academy of Sciences* **114**, E4149 (2017).
- <sup>107</sup>K. Kaneko, “Lyapunov analysis and information flow in coupled map lattices”, *Physica D: Nonlinear Phenomena* **23**, 436–447 (1986).

- <sup>108</sup>K. Kaneko, “Pattern dynamics in spatiotemporal chaos: Pattern selection, diffusion of defect and pattern competition intermittency”, *Physica D: Nonlinear Phenomena* **34**, 1–41 (1989).
- <sup>109</sup>K. Kaneko, “Overview of coupled map lattices”, *Chaos: An Interdisciplinary Journal of Nonlinear Science* **2**, 279–282 (1992).
- <sup>110</sup>K. Kaneko, *Life: An Introduction to Complex Systems Biology*, 1st ed. (Springer-Verlag Berlin Heidelberg, 2006).
- <sup>111</sup>H. B. Kaplan and E. P. Greenberg, “Diffusion of autoinducer is involved in regulation of the *Vibrio fischeri* luminescence system”, *Journal of bacteriology* **163**, 1210–1214 (1985).
- <sup>112</sup>J. Kari, “Reversibility and surjectivity problems of cellular automata”, *Journal of Computer and System Sciences* **48**, 149–182 (1994).
- <sup>113</sup>D. Karig, K. M. Martini, T. Lu, N. A. DeLateur, N. Goldenfeld, and R. Weiss, “Stochastic Turing patterns in a synthetic bacterial population”, *Proceedings of the National Academy of Sciences* **115**, 6572 (2018).
- <sup>114</sup>A. Kicheva, M. Cohen, and J. Briscoe, “Developmental Pattern Formation: Insights from Physics and Biology”, *Science* **338**, 210 (2012).
- <sup>115</sup>H. Kitano, “Towards a theory of biological robustness”, *Molecular Systems Biology* **3**, 137 (2007).
- <sup>116</sup>S. Kondo and R. Asai, “A reaction–diffusion wave on the skin of the marine angelfish *Pomacanthus*”, *Nature* **376**, 765–768 (1995).
- <sup>117</sup>S. Kondo and T. Miura, “Reaction-Diffusion Model as a Framework for Understanding Biological Pattern Formation”, *Science* **329**, 1616–1620 (2010).
- <sup>118</sup>W. Krauth, *Statistical Mechanics: Algorithms and Computations* (Oxford University Press, 2006).
- <sup>119</sup>A. D. Lander, Q. Nie, and F. Y. M. Wan, “Do Morphogen Gradients Arise by Diffusion?”, *Developmental Cell* **2**, 785–796 (2002).
- <sup>120</sup>C. G. Langton, “Computation at the edge of chaos: Phase transitions and emergent computation”, *Physica D: Nonlinear Phenomena* **42**, 12–37 (1990).
- <sup>121</sup>G. Lebon, D. Jou, and J. Casas-Vázquez, *Understanding Non-equilibrium Thermodynamics: Foundations, Applications, Frontiers* (Springer-Verlag Berlin Heidelberg, 2008).
- <sup>122</sup>F. Lederer, G. I. Stegeman, D. N. Christodoulides, G. Assanto, M. Segev, and Y. Silberberg, “Discrete solitons in optics”, *Physics Reports* **463**, 1–126 (2008).
- <sup>123</sup>P. Legendre, “Spatial Autocorrelation: Trouble or New Paradigm?”, *Ecology* **74**, 1659–1673 (1993).
- <sup>124</sup>Z. Li, S. Liu, and Q. Yang, “Incoherent Inputs Enhance the Robustness of Biological Oscillators”, *Cell Systems* **5**, 72–81.e4 (2017).

- <sup>125</sup>S. C. Little, G. Tkačik, T. B. Kneeland, E. F. Wieschaus, and T. Gregor, “The Formation of the Bicoid Morphogen Gradient Requires Protein Movement from Anteriorly Localized mRNA”, *PLOS Biology* **9**, e1000596 (2011).
- <sup>126</sup>C. Liu, X. Fu, L. Liu, X. Ren, C. K. L. Chau, S. Li, L. Xiang, H. Zeng, G. Chen, L.-H. Tang, P. Lenz, X. Cui, W. Huang, T. Hwa, and J.-D. Huang, “Sequential Establishment of Stripe Patterns in an Expanding Cell Population”, *Science* **334**, 238 (2011).
- <sup>127</sup>J. Liu, R. Martinez-Corral, A. Prindle, D.-y. D. Lee, J. Larkin, M. Gabalda-Sagarra, J. Garcia-Ojalvo, and G. M. Süel, “Coupling between distant biofilms and emergence of nutrient time-sharing”, *Science* **356**, 638 (2017).
- <sup>128</sup>J. Liu, A. Prindle, J. Humphries, M. Gabalda-Sagarra, M. Asally, D.-y. D. Lee, S. Ly, J. Garcia-Ojalvo, and G. M. Süel, “Metabolic co-dependence gives rise to collective oscillations within biofilms”, *Nature* **523**, 550 (2015).
- <sup>129</sup>M. Loose, E. Fischer-Friedrich, J. Ries, K. Kruse, and P. Schwille, “Spatial Regulators for Bacterial Cell Division Self-Organize into Surface Waves in Vitro”, *Science* **320**, 789 (2008).
- <sup>130</sup>D. K. Lubensky, M. W. Pennington, B. I. Shraiman, and N. E. Baker, “A dynamical model of ommatidial crystal formation”, *Proceedings of the National Academy of Sciences* **108**, 11145 (2011).
- <sup>131</sup>W. Ma, L. Lai, Q. Ouyang, and C. Tang, “Robustness and modular design of the *Drosophila* segment polarity network”, *Molecular Systems Biology* **2**, 70 (2006).
- <sup>132</sup>W. Ma, A. Trusina, H. El-Samad, W. A. Lim, and C. Tang, “Defining Network Topologies that Can Achieve Biochemical Adaptation”, *Cell* **138**, 760–773 (2009).
- <sup>133</sup>J. Macia, R. Manzoni, N. Conde, A. Urrios, E. de Nadal, R. Solé, and F. Posas, “Implementation of Complex Biological Logic Circuits Using Spatially Distributed Multicellular Consortia”, *PLOS Computational Biology* **12**, e1004685 (2016).
- <sup>134</sup>P. Mahalwar, B. Walderich, A. P. Singh, and C. Nüsslein-Volhard, “Local reorganization of xanthophores fine-tunes and colors the striped pattern of zebrafish”, *Science* **345**, 1362 (2014).
- <sup>135</sup>T. Maire and H. Youk, “Molecular-Level Tuning of Cellular Autonomy Controls the Collective Behaviors of Cell Populations”, *Cell Systems* **1**, 349–360 (2015a).
- <sup>136</sup>T. Maire and H. Youk, “A Collective Path toward Regeneration”, *Cell* **161**, 195–196 (2015b).
- <sup>137</sup>T. Mammoto and D. E. Ingber, “Mechanical control of tissue and organ development”, *Development* **137**, 1407 (2010).
- <sup>138</sup>Manu, S. Surkova, A. V. Spirov, V. V. Gursky, H. Janssens, A.-R. Kim, O. Radulescu, C. E. Vanario-Alonso, D. H. Sharp, M. Samsonova, and J. Reinitz, “Canalization of Gene Expression and Domain Shifts in the *Drosophila* Blastoderm by Dynamical Attractors”, *PLOS Computational Biology* **5**, e1000303 (2009).

- <sup>139</sup>L. Manukyan, S. A. Montandon, A. Fofonjka, S. Smirnov, and M. C. Milinkovitch, “A living mesoscopic cellular automaton made of skin scales”, *Nature* **544**, 173 (2017).
- <sup>140</sup>L. Marcon, X. Diego, J. Sharpe, and P. Müller, “High-throughput mathematical analysis identifies Turing networks for patterning with equally diffusing signals”, *eLife* **5**, e14022 (2016).
- <sup>141</sup>L. Marcon and J. Sharpe, “Turing patterns in development: what about the horse part?”, *Current Opinion in Genetics & Development* **22**, 578–584 (2012).
- <sup>142</sup>R. Martinez-Corral, J. Liu, G. M. Süel, and J. Garcia-Ojalvo, “Bistable emergence of oscillations in growing *Bacillus subtilis* biofilms”, *Proceedings of the National Academy of Sciences* **115**, E8333 (2018).
- <sup>143</sup>N. McDowell, J. Gurdon, and D. J. Grainger, “Formation of a functional morphogen gradient by a passive process in tissue from the early *Xenopus* embryo”, *The International journal of developmental biology* **45**, 199–207 (2001).
- <sup>144</sup>P. Mehta, S. Goyal, T. Long, B. L. Bassler, and N. S. Wingreen, “Information processing and signal integration in bacterial quorum sensing”, *Molecular Systems Biology* **5**, 325 (2009).
- <sup>145</sup>P. Mehta, M. Bukov, C.-H. Wang, A. G. R. Day, C. Richardson, C. K. Fisher, and D. J. Schwab, “A high-bias, low-variance introduction to Machine Learning for physicists”, *Physics Reports* **810**, 1–124 (2019).
- <sup>146</sup>J. Menéndez, A. Pérez-Garijo, M. Calleja, and G. Morata, “A tumor-suppressing mechanism in *Drosophila* involving cell competition and the Hippo pathway”, *Proceedings of the National Academy of Sciences* **107**, 14651 (2010).
- <sup>147</sup>R. M. H. Merks and J. A. Glazier, “A cell-centered approach to developmental biology”, *Physica A: Statistical Mechanics and its Applications* **352**, 113–130 (2005).
- <sup>148</sup>M. Merle, L. Messio, and J. Mozziconacci, “Turing-like patterns in an asymmetric dynamic Ising model”, *Physical Review E* **100**, 042111 (2019).
- <sup>149</sup>T. Mikeladze-Dvali, M. F. Wernet, D. Pistillo, E. O. Mazzoni, A. A. Teleman, Y.-W. Chen, S. Cohen, and C. Desplan, “The Growth Regulators warts/lats and melted Interact in a Bistable Loop to Specify Opposite Fates in *Drosophila* R8 Photoreceptors”, *Cell* **122**, 775–787 (2005).
- <sup>150</sup>M. B. Miller and B. L. Bassler, “Quorum Sensing in Bacteria”, *Annual Review of Microbiology* **55**, 165–199 (2001).
- <sup>151</sup>M. Mitchell, P. T. Hraber, and J. P. Crutchfield, “Revisiting the Edge of Chaos: Evolving Cellular Automata to Perform Computations”, *Complex Systems* **7**, 89–130 (1993).
- <sup>152</sup>S. De Monte and F. d’Ovidio, “Dynamics of order parameters for globally coupled oscillators”, *Europhysics Letters (EPL)* **58**, 21–27 (2002).
- <sup>153</sup>S. De Monte, F. d’Ovidio, and E. Mosekilde, “Coherent Regimes of Globally Coupled Dynamical Systems”, *Physical Review Letters* **90**, 054102 (2003).

- <sup>154</sup>S. De Monte, F. Ovidio, S. Danø, and P. G. Sørensen, “Dynamical quorum sensing: Population density encoded in cellular dynamics”, *Proceedings of the National Academy of Sciences* **104**, 18377 (2007).
- <sup>155</sup>T. Mora and W. Bialek, “Are Biological Systems Poised at Criticality?”, *Journal of Statistical Physics* **144**, 268–302 (2011).
- <sup>156</sup>P. A. P. Moran, “Notes on Continuous Stochastic Phenomena”, *Biometrika* **37**, 17–23 (1950).
- <sup>157</sup>L. G. Morelli, K. Uriu, S. Ares, and A. C. Oates, “Computational Approaches to Developmental Patterning”, *Science* **336**, 187 (2012).
- <sup>158</sup>P. Müller, K. W. Rogers, B. M. Jordan, J. S. Lee, D. Robson, S. Ramanathan, and A. F. Schier, “Differential Diffusivity of Nodal and Lefty Underlies a Reaction-Diffusion Patterning System”, *Science* **336**, 721 (2012).
- <sup>159</sup>J. D. Murray, “Parameter space for turing instability in reaction diffusion mechanisms: A comparison of models”, *Journal of Theoretical Biology* **98**, 143–163 (1982).
- <sup>160</sup>J. Murray and M. Myerscough, “Pigmentation pattern formation on snakes”, *Journal of Theoretical Biology* **149**, 339–360 (1991).
- <sup>161</sup>J. D. Murray, *Mathematical Biology. II: Spatial models and biomedical applications*, 3rd, Vol. II, Interdisciplinary applied mathematics (Springer, New York, 2003).
- <sup>162</sup>J. Nagumo, S. Arimoto, and S. Yoshizawa, “An Active Pulse Transmission Line Simulating Nerve Axon”, *Proceedings of the IRE* **50**, 2061–2070 (1962).
- <sup>163</sup>H. Nakao and A. S. Mikhailov, “Turing patterns in network-organized activator–inhibitor systems”, *Nature Physics* **6**, 544 (2010).
- <sup>164</sup>K. H. Nealson, T. Platt, and J. W. Hastings, “Cellular control of the synthesis and activity of the bacterial luminescent system”, *Journal of bacteriology* **104**, 313–322 (1970).
- <sup>165</sup>G. Nicolis and I. Prigogine, *Self-Organization in Nonequilibrium Systems: From Dissipative Structures to Order through Fluctuations* (Wiley, 1977).
- <sup>166</sup>H. Niederholtmeyer, Z. Z. Sun, Y. Hori, E. Yeung, A. Verpoorte, R. M. Murray, and S. J. Maerkl, “Rapid cell-free forward engineering of novel genetic ring oscillators”, *eLife* **4**, e09771 (2015).
- <sup>167</sup>A. C. Oates, L. G. Morelli, and S. Ares, “Patterning embryos with oscillations: structure, function and dynamics of the vertebrate segmentation clock”, *Development* **139**, 625 (2012).
- <sup>168</sup>E. P. Olimpio, “Topics on Statistical Mechanics of Cells”, MSc Thesis (Delft University of Technology, 2016).
- <sup>169</sup>E. P. Olimpio, Y. Dang, and H. Youk, “Statistical Dynamics of Spatial-Order Formation by Communicating Cells”, *iScience* **2**, 27–40 (2018).
- <sup>170</sup>M. Osswald, E. Jung, F. Sahn, G. Solecki, V. Venkataramani, J. Blaes, S. Weil, H. Horstmann, B. Wiestler, M. Syed, L. Huang, M. Ratliff, K. Karimian Jazi, F. T. Kurz, T. Schmenger, D.

- Lemke, M. Gömmel, M. Pauli, Y. Liao, P. Häring, S. Pusch, V. Herl, C. Steinhäuser, D. Kronic, M. Jarahian, H. Miletic, A. S. Berghoff, O. Griesbeck, G. Kalamakis, O. Garaschuk, M. Preusser, S. Weiss, H. Liu, S. Heiland, M. Platten, P. E. Huber, T. Kuner, A. von Deimling, W. Wick, and F. Winkler, “Brain tumour cells interconnect to a functional and resistant network”, *Nature* **528**, 93 (2015).
- <sup>171</sup>G. F. Oster, J. D. Murray, and A. K. Harris, “Mechanical aspects of mesenchymal morphogenesis”, *Development* **78**, 83–125 (1983).
- <sup>172</sup>A. Oyler-Yaniv, J. Oyler-Yaniv, B. M. Whitlock, Z. Liu, R. N. Germain, M. Huse, G. Altan-Bonnet, and O. Krichevsky, “A Tunable Diffusion-Consumption Mechanism of Cytokine Propagation Enables Plasticity in Cell-to-Cell Communication in the Immune System”, *Immunity* **46**, 609–620 (2017).
- <sup>173</sup>K. M. Page, P. K. Maini, and N. A. M. Monk, “Complex pattern formation in reaction–diffusion systems with spatially varying parameters”, *Physica D: Nonlinear Phenomena* **202**, 95–115 (2005).
- <sup>174</sup>A. Pai, Y. Tanouchi, and L. You, “Optimality and robustness in quorum sensing (QS)-mediated regulation of a costly public good enzyme”, *Proceedings of the National Academy of Sciences* **109**, 19810 (2012).
- <sup>175</sup>K. Papenfort and B. L. Bassler, “Quorum sensing signal-response systems in Gram-negative bacteria”, *Nature Reviews Microbiology* **14**, 576 (2016).
- <sup>176</sup>A. C. Paz, S. Javaherian, and A. P. McGuigan, “Tools for micropatterning epithelial cells into microcolonies on transwell filter substrates”, *Lab on a Chip* **11**, 3440–3448 (2011).
- <sup>177</sup>J. P. Pearson, C. Van Delden, and B. H. Iglewski, “Active Efflux and Diffusion Are Involved in Transport of *Pseudomonas aeruginosa* Cell-to-Cell Signals”, *Journal of Bacteriology* **181**, 1203 (1999).
- <sup>178</sup>R. Pepperkok and J. Ellenberg, “High-throughput fluorescence microscopy for systems biology”, *Nature Reviews Molecular Cell Biology* **7**, 690–696 (2006).
- <sup>179</sup>N. Perunov, R. A. Marsland, and J. L. England, “Statistical Physics of Adaptation”, *Physical Review X* **6**, 021036 (2016).
- <sup>180</sup>R. Phillips, “Napoleon Is in Equilibrium”, *Annual Review of Condensed Matter Physics* **6**, 85–111 (2015).
- <sup>181</sup>N. Posnien, C. Hopfen, M. Hilbrant, M. Ramos-Womack, S. Murat, A. Schönauer, S. L. Herbert, M. D. S. Nunes, S. Arif, C. J. Breuker, C. Schlötterer, P. Mitteroecker, and A. P. McGregor, “Evolution of Eye Morphology and Rhodopsin Expression in the *Drosophila melanogaster* Species Subgroup”, *PLOS ONE* **7**, e37346 (2012).
- <sup>182</sup>O. Pourquié, “Vertebrate Segmentation: From Cyclic Gene Networks to Scoliosis”, *Cell* **145**, 650–663 (2011).
- <sup>183</sup>I. Prigogine, *Nobel Lecture: Time, Structure and Fluctuations*, Report (NobelPrize.org, 1977).

- <sup>184</sup>I. Prigogine and I. Stengers, *Order out of chaos: Man's new dialogue with nature* (Bantam New Age Books, 1984).
- <sup>185</sup>A. Prindle, J. Liu, M. Asally, S. Ly, J. Garcia-Ojalvo, and G. M. Süel, "Ion channels enable electrical communication in bacterial communities", *Nature* **527**, 59 (2015).
- <sup>186</sup>P. Prusinkiewicz and A. Lindenmayer, *The Algorithmic Beauty of Plants* (Springer-Verlag New York Inc., 1996).
- <sup>187</sup>A. Raj, P. van den Bogaard, S. A. Rifkin, A. van Oudenaarden, and S. Tyagi, "Imaging individual mRNA molecules using multiple singly labeled probes", *Nature Methods* **5**, 877 (2008).
- <sup>188</sup>A. Raj and A. van Oudenaarden, "Nature, Nurture, or Chance: Stochastic Gene Expression and Its Consequences", *Cell* **135**, 216–226 (2008).
- <sup>189</sup>F.-A. Ramírez-Weber and T. B. Kornberg, "Cytosomes: Cellular Processes that Project to the Principal Signaling Center in *Drosophila* Imaginal Discs", *Cell* **97**, 599–607 (1999).
- <sup>190</sup>J. Raspopovic, L. Marcon, L. Russo, and J. Sharpe, "Digit patterning is controlled by a Bmp-Sox9-Wnt Turing network modulated by morphogen gradients", *Science* **345**, 566 (2014).
- <sup>191</sup>P. Recho, A. Hallou, and E. Hannezo, "Theory of mechanochemical patterning in biphasic biological tissues", *Proceedings of the National Academy of Sciences* **116**, 5344 (2019).
- <sup>192</sup>P. Rendell, "A Universal Turing Machine in Conway's Game of Life", in 2011 International Conference on High Performance Computing & Simulation (2011), pp. 764–772.
- <sup>193</sup>J. F. Reverey, J.-H. Jeon, H. Bao, M. Leippe, R. Metzler, and C. Selhuber-Unkel, "Superdiffusion dominates intracellular particle motion in the supercrowded cytoplasm of pathogenic *Acanthamoeba castellanii*", *Scientific Reports* **5**, 11690 (2015).
- <sup>194</sup>I. H. Riedel-Kruse, C. Müller, and A. C. Oates, "Synchrony Dynamics During Initiation, Failure, and Rescue of the Segmentation Clock", *Science* **317**, 1911 (2007).
- <sup>195</sup>E. K. Rodriguez, A. Hoger, and A. D. McCulloch, "Stress-dependent finite growth in soft elastic tissues", *Journal of Biomechanics* **27**, 455–467 (1994).
- <sup>196</sup>F. Rosas, A. P. Mediano, M. Ugarte, and J. H. Jensen, "An Information-Theoretic Approach to Self-Organisation: Emergence of Complex Interdependencies in Coupled Dynamical Systems", *Entropy* **20**, 10.3390/e20100793 (2018).
- <sup>197</sup>H. Rouault and V. Hakim, "Different Cell Fates from Cell-Cell Interactions: Core Architectures of Two-Cell Bistable Networks", *Biophysical Journal* **102**, 417–426 (2012).
- <sup>198</sup>M. Ruse, "17. From Organicism to Mechanism-and Halfway Back?", in *Beyond Mechanism: Putting Life Back Into Biology*, edited by A. Henning Brian G.; Scarfe (Lexington Books, 2013), p. 484.
- <sup>199</sup>A. Rustom, R. Saffrich, I. Markovic, P. Walther, and H.-H. Gerdes, "Nanotubular Highways for Intercellular Organelle Transport", *Science* **303**, 1007 (2004).

- <sup>200</sup>W. van Saarloos, “Front propagation into unstable states”, *Physics Reports* **386**, 29–222 (2003).
- <sup>201</sup>F. Sagués, J. M. Sancho, and J. García-Ojalvo, “Spatiotemporal order out of noise”, *Reviews of Modern Physics* **79**, 829–882 (2007).
- <sup>202</sup>A. Sanchez and I. Golding, “Genetic Determinants and Cellular Constraints in Noisy Gene Expression”, *Science* **342**, 1188 (2013).
- <sup>203</sup>J. Santos-Moreno and Y. Schaerli, “Using Synthetic Biology to Engineer Spatial Patterns”, *Advanced Biosystems* **0**, 1800280 (2018).
- <sup>204</sup>Y. Sasai, “Cytosystems dynamics in self-organization of tissue architecture”, *Nature* **493**, 318 (2013).
- <sup>205</sup>S. Sawai, P. A. Thomason, and E. C. Cox, “An autoregulatory circuit for long-range self-organization in *Dictyostelium* cell populations”, *Nature* **433**, 323–326 (2005).
- <sup>206</sup>M. Schienbein and H. Gruler, “Langevin equation, Fokker-Planck equation and cell migration”, *Bulletin of Mathematical Biology* **55**, 585–608 (1993).
- <sup>207</sup>E. Schneidman, M. J. Berry, R. Segev, and W. Bialek, “Weak pairwise correlations imply strongly correlated network states in a neural population”, *Nature* **440**, 1007–1012 (2006).
- <sup>208</sup>N. S. Scholes and M. Isalan, “A three-step framework for programming pattern formation”, *Current Opinion in Chemical Biology* **40**, 1–7 (2017).
- <sup>209</sup>N. S. Scholes, D. Schnoerr, M. Isalan, and M. P. H. Stumpf, “A Comprehensive Network Atlas Reveals That Turing Patterns Are Common but Not Robust”, *Cell Systems* **9**, 243–257.e4 (2019).
- <sup>210</sup>M. Scianna, C. G. Bell, and L. Preziosi, “A review of mathematical models for the formation of vascular networks”, *Journal of Theoretical Biology* **333**, 174–209 (2013).
- <sup>211</sup>X. Serra-Picamal, V. Conte, R. Vincent, E. Anon, D. T. Tambe, E. Bazellieres, J. P. Butler, J. J. Fredberg, and X. Trepat, “Mechanical waves during tissue expansion”, *Nature Physics* **8**, 628 (2012).
- <sup>212</sup>A. E. Sgro, “Who Said To Do That? Understanding Multicellular Decision Making, presented at the Bionanoscience departmental seminar”, Unpublished Work, 2019.
- <sup>213</sup>A. E. Sgro, D. J. Schwab, J. Noorbakhsh, T. Mestler, P. Mehta, and T. Gregor, “From intracellular signaling to population oscillations: bridging size- and time-scales in collective behavior”, *Molecular Systems Biology* **11** (2015).
- <sup>214</sup>C. E. Shannon, “A Mathematical Theory of Communication”, *Bell System Technical Journal* **27**, 379–423 (1948).
- <sup>215</sup>J. Sharpe, “Computer modeling in developmental biology: growing today, essential tomorrow”, *Development* **144**, 4214 (2017).
- <sup>216</sup>D. Sherrington and S. Kirkpatrick, “Solvable Model of a Spin-Glass”, *Physical Review Letters* **35**, 1792–1796 (1975).

- <sup>217</sup>R. Sheth, L. Marcon, M. F. Bastida, M. Junco, L. Quintana, R. Dahn, M. Kmita, J. Sharpe, and M. A. Ros, “Hox Genes Regulate Digit Patterning by Controlling the Wavelength of a Turing-Type Mechanism”, *Science* **338**, 1476 (2012).
- <sup>218</sup>S. Y. Shvartsman, H. S. Wiley, W. M. Deen, and D. A. Lauffenburger, “Spatial Range of Autocrine Signaling: Modeling and Computational Analysis”, *Biophysical Journal* **81**, 1854–1867 (2001).
- <sup>219</sup>A. E. Shyer, T. R. Huycke, C. Lee, L. Mahadevan, and C. J. Tabin, “Bending Gradients: How the Intestinal Stem Cell Gets Its Home”, *Cell* **161**, 569–580 (2015).
- <sup>220</sup>S. Sick, S. Reinker, J. Timmer, and T. Schlake, “WNT and DKK Determine Hair Follicle Spacing Through a Reaction-Diffusion Mechanism”, *Science* **314**, 1447 (2006).
- <sup>221</sup>C. I. E. Smith, R. Holmdahl, O. Kämpe, and K. Kärre, *Scientific Background: Discovery of cancer therapy by inhibition of negative immune regulation*, Press Release 5 (2018).
- <sup>222</sup>B. Snijder and L. Pelkmans, “Origins of regulated cell-to-cell variability”, *Nature Reviews Molecular Cell Biology* **12**, 119 (2011).
- <sup>223</sup>T. R. Sokolowski, T. Erdmann, and P. R. ten Wolde, “Mutual Repression Enhances the Steepness and Precision of Gene Expression Boundaries”, *PLOS Computational Biology* **8**, e1002654 (2012).
- <sup>224</sup>K. F. Sonnen, V. M. Lauschke, J. Uraji, H. J. Falk, Y. Petersen, M. C. Funk, M. Beaupeux, P. François, C. A. Merten, and A. Aulehla, “Modulation of Phase Shift between Wnt and Notch Signaling Oscillations Controls Mesoderm Segmentation”, *Cell* **172**, 1079–1090.e12 (2018).
- <sup>225</sup>M. B. Sporn and G. J. Todaro, “Autocrine Secretion and Malignant Transformation of Cells”, *New England Journal of Medicine* **303**, 878–880 (1980).
- <sup>226</sup>M. S. Steinberg, “Differential adhesion in morphogenesis: a modern view”, *Current Opinion in Genetics & Development* **17**, 281–286 (2007).
- <sup>227</sup>O. Stempler, A. K. Baidya, S. Bhattacharya, G. B. Malli Mohan, E. Tzipilevich, L. Sinai, G. Mamou, and S. Ben-Yehuda, “Interspecies nutrient extraction and toxin delivery between bacteria”, *Nature Communications* **8**, 315 (2017).
- <sup>228</sup>C. Stokes, D. Lauffenburger, and S. Williams, “Migration of individual microvessel endothelial cells: stochastic model and parameter measurement”, *Journal of Cell Science* **99**, 419–430 (1991).
- <sup>229</sup>J. Stricker, S. Cookson, M. R. Bennett, W. H. Mather, L. S. Tsimring, and J. Hasty, “A fast, robust and tunable synthetic gene oscillator”, *Nature* **456**, 516 (2008).
- <sup>230</sup>S. A. Stricker, “Comparative Biology of Calcium Signaling during Fertilization and Egg Activation in Animals”, *Developmental Biology* **211**, 157–176 (1999).
- <sup>231</sup>S. H. Strogatz, *Nonlinear Dynamics and Chaos: With Applications to Physics, Biology, Chemistry, and Engineering*, 1st ed. (Addison-Wesley, Reading, Massachusetts, USA, 1994).

- <sup>232</sup>S. H. Strogatz, “From Kuramoto to Crawford: exploring the onset of synchronization in populations of coupled oscillators”, *Physica D: Nonlinear Phenomena* **143**, 1–20 (2000).
- <sup>233</sup>A. Szabó and R. M. Merks, “Cellular Potts Modeling of Tumor Growth, Tumor Invasion, and Tumor Evolution”, *Frontiers in Oncology* **3**, 1–12 (2013).
- <sup>234</sup>A. M. Tayar, E. Karzbrun, V. Noireaux, and R. H. Bar-Ziv, “Synchrony and pattern formation of coupled genetic oscillators on a chip of artificial cells”, *Proceedings of the National Academy of Sciences* **114**, 11609 (2017).
- <sup>235</sup>O. Tchernyshyov and G.-W. Chern, “Spin-Lattice Coupling in Frustrated Antiferromagnets”, in *Introduction to Frustrated Magnetism: Materials, Experiments, Theory*, edited by C. Lacroix, P. Mendels, and F. Mila (Springer Berlin Heidelberg, Berlin, Heidelberg, 2011), pp. 269–291.
- <sup>236</sup>A. A. Teleman and S. M. Cohen, “Dpp Gradient Formation in the *Drosophila* Wing Imaginal Disc”, *Cell* **103**, 971–980 (2000).
- <sup>237</sup>D. Thalmeier, J. Halatek, and E. Frey, “Geometry-induced protein pattern formation”, *Proceedings of the National Academy of Sciences* **113**, 548 (2016).
- <sup>238</sup>M. Théry, “Micropatterning as a tool to decipher cell morphogenesis and functions”, *Journal of Cell Science* **123**, 4201 (2010).
- <sup>239</sup>R. Thomas and R. D’Ari, *Biological Feedback* (CRC Press, Inc., 1990).
- <sup>240</sup>D. W. Thompson, *On growth and form*, 2nd (Dover Publications; Revised edition, Mineola, New York, 1942).
- <sup>241</sup>K. Thurley, L. F. Wu, and S. J. Altschuler, “Modeling Cell-to-Cell Communication Networks Using Response-Time Distributions”, *Cell Systems* **6**, 355–367.e5 (2018).
- <sup>242</sup>G. Tkačik, T. Gregor, and W. Bialek, “The Role of Input Noise in Transcriptional Regulation”, *PLOS ONE* **3**, e2774 (2008).
- <sup>243</sup>G. Tkačik and A. M. Walczak, “Information transmission in genetic regulatory networks: a review”, *Journal of Physics: Condensed Matter* **23**, 153102 (2011).
- <sup>244</sup>I. M. Tolić-Nørrelykke, E.-L. Munteanu, G. Thon, L. Oddershede, and K. Berg-Sørensen, “Anomalous Diffusion in Living Yeast Cells”, *Physical Review Letters* **93**, 078102 (2004).
- <sup>245</sup>T. Tomka, D. Iber, and M. Boaretto, “Travelling waves in somitogenesis: Collective cellular properties emerge from time-delayed juxtacrine oscillation coupling”, *Progress in Biophysics and Molecular Biology* **137**, 76–87 (2018).
- <sup>246</sup>J. Toner and Y. Tu, “Flocks, herds, and schools: A quantitative theory of flocking”, *Physical Review E* **58**, 4828–4858 (1998).
- <sup>247</sup>F. Tostevin and P. R. ten Wolde, “Mutual Information between Input and Output Trajectories of Biochemical Networks”, *Physical Review Letters* **102**, 218101 (2009).
- <sup>248</sup>F. Tostevin and P. R. ten Wolde, “Mutual information in time-varying biochemical systems”, *Physical Review E* **81**, 061917 (2010).

- <sup>249</sup>N. B. Trunnell, A. C. Poon, S. Y. Kim, and J. E. Ferrell, “Ultrasensitivity in the Regulation of Cdc25C by Cdk1”, *Molecular Cell* **41**, 263–274 (2011).
- <sup>250</sup>M. Tsachaki and S. G. Sprecher, “Genetic and developmental mechanisms underlying the formation of the *Drosophila* compound eye”, *Developmental Dynamics* **241**, 40–56 (2012).
- <sup>251</sup>T. Y.-C. Tsai, Y. S. Choi, W. Ma, J. R. Pomerening, C. Tang, and J. E. Ferrell, “Robust, Tunable Biological Oscillations from Interlinked Positive and Negative Feedback Loops”, *Science* **321**, 126 (2008).
- <sup>252</sup>C. D. Tsiairis and A. Aulehla, “Self-Organization of Embryonic Genetic Oscillators into Spatiotemporal Wave Patterns”, *Cell* **164**, 656–667 (2016).
- <sup>253</sup>J. J. Tyson, K. C. Chen, and B. Novák, “Sniffers, buzzers, toggles and blinkers: dynamics of regulatory and signaling pathways in the cell”, *Current Opinion in Cell Biology* **15**, 221–231 (2003).
- <sup>254</sup>J. J. Tyson and B. Novák, “Functional Motifs in Biochemical Reaction Networks”, *Annual Review of Physical Chemistry* **61**, 219–240 (2010).
- <sup>255</sup>T. Umeda and K. Inouye, “Cell sorting by differential cell motility: a model for pattern formation in *Dictyostelium*”, *Journal of Theoretical Biology* **226**, 215–224 (2004).
- <sup>256</sup>J. Urías, R. Rechtman, and A. Enciso, “Sensitive dependence on initial conditions for cellular automata”, *Chaos: An Interdisciplinary Journal of Nonlinear Science* **7**, 688–693 (1997).
- <sup>257</sup>T. Vicsek, A. Czirók, E. Ben-Jacob, I. Cohen, and O. Shochet, “Novel Type of Phase Transition in a System of Self-Driven Particles”, *Physical Review Letters* **75**, 1226–1229 (1995).
- <sup>258</sup>V. Vogel and M. Sheetz, “Local force and geometry sensing regulate cell functions”, *Nature Reviews Molecular Cell Biology* **7**, 265–275 (2006).
- <sup>259</sup>A. Volkening and B. Sandstede, “Iridophores as a source of robustness in zebrafish stripes and variability in Danio patterns”, *Nature Communications* **9**, 3231 (2018).
- <sup>260</sup>G. von Dassow, E. Meir, E. M. Munro, and G. M. Odell, “The segment polarity network is a robust developmental module”, *Nature* **406**, 188 (2000).
- <sup>261</sup>M. Watanabe and S. Kondo, “Is pigment patterning in fish skin determined by the Turing mechanism?”, *Trends in Genetics* **31**, 88–96 (2015).
- <sup>262</sup>J. R. Weimar and J.-P. Boon, “Class of cellular automata for reaction-diffusion systems”, *Physical Review E* **49**, 1749–1752 (1994).
- <sup>263</sup>O. D. Weiner, W. A. Marganski, L. F. Wu, S. J. Altschuler, and M. W. Kirschner, “An Actin-Based Wave Generator Organizes Cell Motility”, *PLOS Biology* **5**, e221 (2007).
- <sup>264</sup>P. R. Ten Wolde, N. B. Becker, T. E. Ouldridge, and A. Mugler, “Fundamental Limits to Cellular Sensing”, *Journal of Statistical Physics* **162**, 1395–1424 (2016).

- <sup>265</sup>S. Wolfram, “Statistical mechanics of cellular automata”, *Reviews of Modern Physics* **55**, 601–644 (1983).
- <sup>266</sup>S. Wolfram, “Cellular automata as models of complexity”, *Nature* **311**, 419–424 (1984).
- <sup>267</sup>S. Wolfram, *A New Kind of Science* (Wolfram Media, Champaign, IL, 2002).
- <sup>268</sup>L. Wolpert, “Chapter 6: Positional Information and Pattern Formation”, in *Current Topics in Developmental Biology*, Vol. 6, edited by A. A. Moscona and A. Monroy (Academic Press, 1971), pp. 183–224.
- <sup>269</sup>F. Wu, B. G. C. van Schie, J. E. Keymer, and C. Dekker, “Symmetry and scale orient Min protein patterns in shaped bacterial sculptures”, *Nature Nanotechnology* **10**, 719 (2015).
- <sup>270</sup>M. A. Wyczalkowski, Z. Chen, B. A. Filas, V. D. Varner, and L. A. Taber, “Computational models for mechanics of morphogenesis”, *Birth Defects Research Part C: Embryo Today: Reviews* **96**, 132–152 (2012).
- <sup>271</sup>H. Xu, S. O. Skinner, A. M. Sokac, and I. Golding, “Stochastic Kinetics of Nascent RNA”, *Physical Review Letters* **117**, 128101 (2016).
- <sup>272</sup>C.-h. Yang, J. D. Axelrod, and M. A. Simon, “Regulation of Frizzled by Fat-like Cadherins during Planar Polarity Signaling in the *Drosophila* Compound Eye”, *Cell* **108**, 675–688 (2002).
- <sup>273</sup>H. W. Yang, S. R. Collins, and T. Meyer, “Locally excitable Cdc42 signals steer cells during chemotaxis”, *Nature Cell Biology* **18**, 191 (2015).
- <sup>274</sup>L. You, R. S. Cox, R. Weiss, and F. H. Arnold, “Programmed population control by cell–cell communication and regulated killing”, *Nature* **428**, 868–871 (2004).
- <sup>275</sup>H. Youk and W. A. Lim, “Secreting and Sensing the Same Molecule Allows Cells to Achieve Versatile Social Behaviors”, *Science* **343**, 1242782 (2014a).
- <sup>276</sup>H. Youk and W. A. Lim, “Sending Mixed Messages for Cell Population Control”, *Cell* **158**, 973–975 (2014b).
- <sup>277</sup>G. Young, M. Demir, H. Salman, G. B. Ermentrout, and J. E. Rubin, “Interactions of solitary pulses of *E. coli* in a one-dimensional nutrient gradient”, *Physica D: Nonlinear Phenomena* **395**, 24–36 (2019).
- <sup>278</sup>D. Y. Zhang and G. Seelig, “Dynamic DNA nanotechnology using strand-displacement reactions”, *Nature Chemistry* **3**, 103–113 (2011).
- <sup>279</sup>M. M. Zheng, B. Shao, and Q. Ouyang, “Identifying network topologies that can generate turing pattern”, *Journal of Theoretical Biology* **408**, 88–96 (2016).

# CURRICULUM VITÆ

## Yiteng Dang

- 2015-2019 **Doctor of Philosophy (PhD)**  
Department of Bionanoscience  
Delft University of Technology (TU Delft), Delft, The Netherlands  
*Co-promotor and daily supervisor:* Dr. Hyun Youk  
*Promotor:* Prof. dr. Marileen Dogterom
- 2014-2015 **Master of Science (MSc)**  
International Centre for Fundamental Physics (ICFP) Masters Program  
Quantum Physics track, second year (M2)  
École Normale Supérieure, Paris, France
- 2013-2014 **Master of Advanced Study (MASt)**, with *Distinction*  
Part III of the Mathematical Tripos  
Department of Applied Mathematics and Theoretical Physics  
University of Cambridge, Cambridge, United Kingdom
- 2010-2013 **Bachelor of Science (BSc)**, *cum laude*  
Physics & Mathematics, double degree  
Utrecht University, Utrecht, The Netherlands
- 2004-2010 Secondary School (VWO Gymnasium)  
Marnix College, Ede, The Netherlands
- 13-01-1993 Born in Wuxi, China

# LIST OF PUBLICATIONS

2. **Dang, Y.**, Grundel, D. A. J. & Youk, H., Ways for cells to communicate with diffusible molecules that yield dynamic spatial patterns, *Cell Systems* **10**, 82–98.e7 (2020).
1. Olimpio, E. P.\* , **Dang, Y.\*** & Youk, H. Statistical Dynamics of Spatial-Order Formation by Communicating Cells. *iScience* **2**, 27–40 (2018) (\*co-first authors).