UNIVERSITY OF CALIFORNIA,
IRVINE

Multiscale Modeling of the Epithelial-Mesenchymal Transition

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Mathematics

by

Catherine Ha Ta

Dissertation Committee:
Professor Qing Nie, Chair
Professor Jun Allard
Professor Frederic Yui-Ming Wan

2017

DEDICATION

To my parents, for their unfaltering belief in me

TABLE OF CONTENTS

LIST OF FIGURES

Page

LIST OF TABLES

ACKNOWLEDGMENTS

The text of this dissertation is a reprint of the material as it appears in the Journal of Computational Physics, PLOS Computational Biology, and Discrete and Continuous Dynamical Systems - Series B. The co-author, Qing Nie, listed in all three publications, directed and supervised research which forms the basis for the dissertation.

I would like to express my most sincere gratitude to my committee chair, Professor Qing Nie, for his continuous support of my PhD study and related research, for his patience, immense knowledge, and belief in me. He was constantly challenging and pushing me to outperform myself. I would not have been able to complete this thesis without his constant guidance and encouragement.

I am extremely grateful to the rest of my committee members, Professor Jun Allard and Professor Frederic Y. M. Wan, for their insightful comments, advising, and encouragement.

I thank the PLOS Computational Biology for permission to include Chapter 2 of my dissertation, Discrete and Continuous Dynamical Systems - Series B for permission to include Chapter 3 of the dissertation, and the Journal of Computational Physics for permission to include Chapter 5 of my dissertation. Financial support was provided by NIH grants R01GM107264, R01NS095355, P50GM76516, and R56AR064532, and the NSF grants DMS1161621 and DMS1562176,

I thank Dr. Tian Hong and Dr. Kazuhide Watanabe, the main authors of the PLOS Computational Biology publication, for conceiving the concept behind the publication and for conducting all the experiments. I also want to thank Alvaro Villarreal-Ponce for his participation in these experiments. In addition, I want to take this opportunity to express my deepest appreciation to Dr. Tian Hong, for inspiring me and helping me deepen my existing understanding of the biological components that serve as the background for my thesis. He was always patient and attentive in his responses to my numerous questions, whether it is 2 in the afternoon or 3 in the morning.

Special thanks go to Dr. Xing Dai and Chris Rackauckas for many stimulating discussions, which helped make much of this research possible.

Last but not least, I would like to thank my parents for always believing in me and supporting me morally and financially during my pursuit of a higher education. I also want to express my appreciation for my partner Luke and my two kitties for being my constant companions and for propping me up whenever I am down. Their love and support gave me the strength and courage to face any difficulty that life presents.

CURRICULUM VITAE

Catherine Ha Ta

EDUCATION

Doctor of Philosophy in Mathematics — **Apr 2017**
University of California, Irvine — *Irvine, California*

Master of Science in Mathematics — **Jun 2013**
University of California, Irvine — *Irvine, California*

Bachelor of Science in Mathematics — **Jun 2011**
University of California, Irvine — *Irvine, California*

RESEARCH EXPERIENCE

Graduate Research Assistant — **Sep 2011– Apr 2017**
University of California, Irvine — *Irvine, California*

TEACHING EXPERIENCE

Teaching Assistant — **Sep 2013–Dec 2016**
University of California, Irvine — *Irvine, California*

PUBLIC PRESENTATIONS

Tsukuba-UCI Science Partnership Conference — **Dec 9, 2015**
Irvine, California

SIAM Conference on the Life Sciences — **Jul 11, 2016**
Boston, Massachusetts

REFEREED JOURNAL PUBLICATIONS

An Integration Factor Method for Stochastic and Stiff Reaction-Diffusion Systems **2015**
Journal of Computational Physics

An Ovol2-Zeb1 Mutual Inhibitory Circuit Governs Bidirectional and Multi-step Transition between Epithelial and Mesenchymal States **2015**
PLOS Computational Biology

Controlling stochasticity in epithelial-mesenchymal transition through multiple intermediate cellular states **2016**
Discrete and Continuous Dynamical Systems - Series B

ABSTRACT OF THE DISSERTATION

Multiscale Modeling of the Epithelial-Mesenchymal Transition

By

Catherine Ha Ta

Doctor of Philosophy in Mathematics

University of California, Irvine, 2017

Professor Qing Nie, Chair

Epithelial-mesenchymal transition (EMT) is an instance of cellular plasticity that plays critical roles in development, regeneration and cancer progression. While many regulatory elements have been identified to induce EMT, the complex process underlying such cellular plasticity remains poorly understood. Utilizing a systems biology approach integrating modeling and experiments, we found multiple intermediate states contributing to EMT and that the robustness of the transitions is modulated by transcriptional factor Ovol2. In particular, we observed that adding the mutual inhibition relationship between Ovol2 and EMT inducer Zeb1 generates a novel four-state system consisting of two distinct intermediate phenotypes that differ in differentiation propensities and are favored in different environmental conditions. These intermediate states correspond to various forms of stem-like cells in the EMT system, but the function of the multi-step transition or the multiple stem cell phenotypes is unclear. Here, we used mathematical models to show that multiple intermediate phenotypes in the EMT system help to attenuate the overall fluctuations of the cell population in terms of phenotypic compositions, thereby stabilizing a heterogeneous cell population in the EMT spectrum. We found that the ability of the system to attenuate noise on the intermediate states depends on the number of intermediate states, indicating the stem-cell population is more stable when it has more sub-states. We then attempted to bridge the gap between discrete and continuum modeling of the EMT system by incorporating the EMT

core regulatory network into our heterogeneous cell population dynamics model to create a multiscale EMT model. Our model can capture the larger-scale population growth dynamics while acknowledging the intracellular interactions between proteins within each individual cell. From the two types of noise we introduced into our model, we observed that the differences in the noise design prompted distinctive behaviors in the proliferative capability of our heterogeneous population. Our findings also revealed the challenges encountered when integrating noise into a dynamic EMT model such as the multiscale model and the complex role noise plays in modulating the different phenotypic fractions of the population. Lastly, we present a class of semi-implicit integration factor methods and demonstrate its good accuracy, efficiency, and stability properties compared to existing methods. This new class of methods, which is easy to implement, will have broader applications in solving stochastic reaction-diffusion equations arising from models in biology and physical sciences.

Chapter 1

Introduction

1.1 Epithelial-Mesenchymal transition (EMT)

Epithelial-mesenchymal transition (EMT) is an extreme form of cellular plasticity that is involved in morphogenesis, tissue regeneration and cancer progression. In addition, EMT has been shown to promote stem cell properties, as differentiated epithelial cells that have undergone a round of EMT gain multipotency and self-renewal capability [12, 31, 78]. During EMT, epithelial cells undergo dramatic changes in cell morphology and behavior to form mesenchymal cells. These changes include loss of cell-to-cell junction, loss of cell polarity and acquisition of migratory and invasive properties [14, 50]. The migratory behavior of newly formed mesenchymal cells is critical for the formation of internal organs during embryonic development [14], and it is also involved in cancer metastasis, which often requires the dissemination of cancerous epithelial cells from the primary tumors and subsequent invasion of distant tissues and organs [78]. After arriving at their destinations, mesenchymal cells sometimes revert to epithelial cells to settle, proliferate, and differentiate, via a process called mesenchymal-epithelial transition (MET), suggesting that EMT is a reversible process [5].

Recent theories and experiments showed that EMT does not give rise to terminal mesenchymal phenotype in some biological scenarios [76, 100, 115]. In other words, intermediate epithelial-mesenchymal phenotype may be generated by partial EMT. In addition, it was hypothesized that partial EMT is responsible for collective migration and invasiveness of cancer cells [76]. With the ability to migrate collectively, cancer cells can maintain some cell-cell adhesion and move in concert, while bypassing many constraints that are imposed on individually migrating cells [24, 47, 112]. These observations and hypothesis strongly suggest that EMT is a multi-step transition, and various degrees of the transition may have distinct physiological outcomes.

There is a growing interest in investigating the implications of partial EMT towards cancer growth and progression. Cells with hybrid phenotype are not only perceived to be apoptosis-resistant, but the generation of cancer stem cells (CSCs) has also been identified with the partial EMT state, posing more challenges for devising efficient cancer treatments [29, 48]. Furthermore, the classifications of CSCs into different categories along the EMT spectrum pose important questions on the varying responses CSCs might have to cancer drugs [4, 67]. Consequentially, effective cancer treatment must be tailored to target different differentiated states within the CSC population.

Previous studies have identified key transcription factors (TFs) and microRNAs (miRNAs) that are involved in the regulation of EMT. In particular, mutual inhibition loops formed between Zeb1 and miR200 [6], and between Snail and miR34a [93] are critical components in the regulatory network [84]. Mathematical modeling suggested that these mutual inhibition loops govern a tristable system, in which cells can be stabilized at an epithelial (E) state, a mesenchymal (M) state, or an intermediate state exhibiting expression of signature genes of both E and M in a variable fashion [76, 100]. The intermediate state identified by these models is proposed to associate with cancer cells that exhibit collective migration during tumorigenesis [76], implicating the clinical relevance of the ternary switch in cell plasticity.

1.2 An Ovol2-Zeb1 mutual inhibitory circuit governs bidirectional and multi-step transition between Epithelial and Mesenchymal states

In recent experimental studies, it was shown that transcription factor Ovol2 restricts EMT by directly inhibiting EMT-inducing factors including Zeb1, and that these regulations are critical for proper morphogenesis and for maintaining epithelial lineages in mammary gland and skin epidermis [61, 108]. However, the precise role of Ovol2 in the context of the well-studied core molecular network that controls EMT dynamics remains to be elucidated. In addition, it is unclear how EMT-inhibiting transcriptional factors like Ovol2 and EMT-promoting transcription factors like Zeb1 interact integratively to regulate the intermediate state.

From our collaboration with Xing Dai's lab, our collaborators Watanabe and Villarreal-Ponce provide new experimental evidence suggesting a direct regulation of Ovol2 by Zeb1, which together with previous reports of Ovol2 inhibition of Zeb1 [61, 91, 108] demonstrates the existence of an Ovol2-Zeb1 mutual inhibition circuit. We then present a mathematical model that includes this new regulation, revealing two, rather than one, intermediate states with distinct propensities to differentiate into E and M states. We show that the Ovol2-Zeb1 mutual inhibition circuit is essential for the existence and robustness of both intermediate states in this model, and experimentally validate a specific prediction of the model, namely that Ovol2 is able to reprogram any given state to an E state. Furthermore, Watanabe and Villarreal-Ponce describe experimental results suggesting that mammary epithelial cell line MCF10A represents one of the intermediate states that exhibit a bidirectional potential to differentiate into both E and M states. Our four-state model of the EMT process is consistent with the observations which show four epigenetic states in multi-step EMT [99], multiple epithelial/mesenchymal stem cell phenotypes [3] and four distinct types of ovarian

cancer cells in the EMT spectrum [42]. In addition, Strauss et al. identify a heterogeneous hybrid EMT stage that is composed of two phenotypically different subpopulations from the analysis of epithelial and mesenchymal traits in ovarian cancer [96]. These results help reinforce the notion of an EMT process with multiple intermediate steady states. Together, our findings uncover a new layer of complexity of the dynamic, multi-step transitions between E and M states and unravel key regulatory mechanisms that control such transitions.

1.3 Controlling stochasticity in EMT through multiple intermediate cellular states

Previous theoretical studies suggested the role of heterogeneity in cell population and its feedback control in stem cell regeneration [62]. However, it is unclear why the epithelial/mesenchymal cell population needs these multiple intermediate states, or why the stem cells in the EMT spectrum need to have several subtypes. In particular, what performance objective does the system evolve to achieve by having multiple intermediate states at the expense of simplicity of cellular phenotypes?

To address these questions, we built and compared a series of mathematical models containing various number of intermediate states in EMT spectrum. By analyzing the behaviors of the system in facing noise in cell populations, we found that multiple intermediate phenotypes in the system help to attenuate the overall fluctuations of the cell population in terms of phenotypic compositions, suggesting their ability to stabilize a heterogeneous cell population. When we compared the models with at least one intermediate state, we found that the number of intermediate states positively correlates with the ability to attenuate noise on the intermediate states, indicating that the stem-cell population is more stable when it has more sub-states. These results suggest a performance objective that the EMT system might

have evolved to achieve by having multiple intermediate phenotypes, and it sheds light on the general principles of heterogeneous stem cell population in term of systems design.

1.4 A multiscale model of a heterogeneous population with cell state transitions in EMT and MET directions

Previous research, including our modeling of the four-state EMT process, employs the EMT regulatory network as the main framework to understand the mechanisms that govern the process and classify the different transition stages of EMT based on the gene expression levels of key proteins [47, 48, 76, 100, 115]. However, this discrete (individual-based) framework does not encompass all the cellular activities that take place in a heterogeneous cell population since the focus is only on the changes that happen within one entity, i.e one cell. On the other hand, continuum (population-based) models are population-centric, thus it tends to overlook the genetic or epigenetic regulatory mechanisms at the individual cell level and cannot describe the EMT process where individual cell effects are important [19]. As a result, hybrid models that unite discrete and population-based approaches are increasingly favored in cancer modeling. The multiscale approach has been successfully adapted in the past to model the E-cadherin-β-Catenin pathway and tumor formation in breast cancer [86, 90]. In this section, we attempted to bridge the gap between the discrete EMT and the continuum models by delivering a cohesive multiscale representation that incorporates the regulatory network into the population dynamics. Hence, the word "multiscale" is used in the title of this thesis as an all-encompassing term that can be understood both as "discrete and continuum" and as "gene regulation and cell population dynamics". The formulation of

the multiscale model enables cell lineage tracing of every cell in the heterogeneous population as well as monitoring the progression of gene expressions per cell. Our model is capable of capturing the larger-scale population growth dynamics while acknowledging the intracellular changes that happen within each individual cell. Essentially, the multiscale model facilitates the study of cell proliferation through tracking the variations in expression levels of different genes that are present in all the cells.

To construct our multiscale model, we considered four different populations corresponding to four different phenotypes in the EMT spectrum. We ascribed stem-like state to the I_1 phenotype and allowed all cells to divide a finite number of times before dying, with the exception of stem cells. We then introduced two different types of noise into our model and observed that the differences in the noise design prompted different behaviors in the proliferative capability of our heterogeneous population. While the first type of noise favors the proliferation of stem cell population I_1, leading to an exponential growth in cell population, the second type of noise shows a diminishing I_1 population, with the initial epithelial population contributing the most to the overall population throughout the different cell cycles being considered. In addition, the potential landscapes of the heterogeneous population generated by the two noises illustrate distinct differentiation propensities of the two intermediate states. Our findings reveal the challenges encountered when integrating noise into a dynamic EMT model such as the multiscale model and the complex role noise plays in modulating the different phenotypic fractions of the population.

1.5 An integration factor method for stochastic and stiff reaction-diffusion systems

Lastly, we shift out attention to the simulation aspect of EMT modeling. We discuss the different numerical techniques with wide-ranging applications to solving systems of stochastic differential equations used to mathematically model complex biological systems.

Complex patterns can be extensively found in nature, from the skin of zebrafish to the disposition of feather buds in chicks and hair follicles in mice. Often, those patterns are created by biochemical reactions along with diffusions of the molecules in a cellular or multi-cellular systems [57]. Such biological systems, which may be described in reaction-diffusion equations, are constantly subjected to stochastic effects such as noises and environmental perturbations. The stochastic effects on the biochemical reactions at the single-cell level can result in heterogeneous responses of cellular populations and influence their behaviors [110]. Previous studies on stochasticity reveal the adaptation of biological systems to noise, which can be characterized by the systems' strategies to combat noise, whether by attenuating or exploiting it [106, 110]. For example, spatial stochastic effects help to either prompt the tight localization of proteins or enhance the response to the directional change of a moving pheromone input, resulting in a more robust cell polarization [60]; and the boundary of gene expression domains is sharpened as a result of gene-switching prompted by intracellular noise [116]. It has become increasingly important to incorporate these stochastic effects into the reaction-diffusion equations for better understanding of biological systems.

One can describe a biological system in terms of the following stochastic reaction-diffusion equations:

$$\frac{\partial U}{\partial t} = a\frac{\partial^2 U}{\partial x^2} + f(U) + g(U)\dot{W}(x,t) \tag{1.1}$$

where $\dot{W}(x,t)$ is a standard two-dimensional Wiener process.

One typical way of solving Eq.(1.1) numerically is to apply central difference first to the diffusion and then use the temporal explicit scheme to solve the subsequent system [33], such as using the explicit Euler method [18] or the two, three, and four-stage explicit Runge Kutta schemes for the system containing additive noise and one-dimensional Wiener process in [23]. Another common approach is using the Galerkin projection of the stochastic partial differential equation (SPDE) and then applying the numerical scheme to a finite-dimensional version of the SPDE. For example, Exponential Time Differencing (ETD) scheme may be applied to the Galerkin projection of the SPDE [44, 45]. One other such example is the Lord and Rougemont scheme, which is derived through Galerkin projection and an integrating factor approach [73]. This scheme is most effective for SPDE with Gevrey regularity, and more improvement may be made on such schemes by taking advantage of the Itô-Taylor expansion [73].

While explicit temporal schemes may be directly implemented for various spatial discretization, including finite element and Galerkin methods [35], to deal with the stability constraint associated with the diffusions, one can treat the diffusion term implicitly, while treating other terms explicitly [33] such as implicit Euler and Crank-Nicolson schemes [18]. Higher order methods [55] can be achieved using Galerkin projection and the linear-implicit versions of strong Taylor schemes [44]. Non-uniform time discretization on Brownian motion can also be obtained for implicit Euler scheme [81].

Stochastic stiffness arises from large differences in the magnitudes of Lyapunov exponents [54], resulting in the presence of different time scales. As in the deterministic case, explicit methods face step-size constraint when used to solve stiff stochastic systems [65]. The time-step constraint can be improved with the modification of the stochastic term by adding more terms from the Itô-Taylor expansion for higher order of accuracy and stability. One of the most well-known schemes stemmed from this construction is the Milstein scheme [54]. Treating the stochastic term implicitly is also one of the popular approaches [8, 37, 80, 107], albeit

computationally expensive. Hence, a class of explicit methods known as Chebyshev methods are derived, which have better mean-square stability than explicit Euler method and are not as computationally expensive as implicit methods [1]. A combination of different numerical schemes into one method can also be seen, such as the case of the Composite Euler scheme [9]. For this scheme, at each temporal step, the stochastic differential equation is either solved by implicit Euler method or semi-implicit Euler method. The Composite method has similar order of convergence 1/2 to the Euler Maruyama method but better stability.

Most of the methods mentioned so far are derived to combat the stochastic stiffness through the improvement of the stochastic term, which can be costly and not as effective if the stiffness only occurs in the deterministic term. In such case, methods that treat the deterministic term implicitly while keeping an explicit treatment of the stochastic term are preferred [8]. Here, we propose a new approach to the problem of stiffness caused by the deterministic term, more specifically the reaction term in Eq.(1.1). The approach is based on the semi-Implicit Integration Factor (IIF) methods [69, 82, 83, 104], which has been found to be effective for the stiff reaction-diffusion equations with better stability constraints imposed on the time steps associated with both reaction and diffusion. In this approach, the time-step constraint for the diffusion term arising due to the inverse of the eigenvalues of the diffusion matrix, which can be large in magnitude, is resolved by treating the linear diffusion term exactly using Integration Factor (IF) methods. Such treatment results in an exponential function of the diffusion term and an integral of the nonlinear reaction term, which is then treated using implicit approach through the Lagrange interpolation to deal with its stiffness. Appropriate choices of approximation schemes lead to decoupling on the treatment for the diffusions and reactions such that one only needs to solve nonlinear systems with the size of the original PDEs. The IIF methods also have exceptional stability properties and its second-order version is absolutely stable. For higher-dimensional problems, the compact IIF (cIIF) method [82] is a great improvement on computational efficiency without altering the stability properties of the IIF methods [83].

In the last section of this thesis, we exploit the simple structure of the IIF methods as well as their desirable stability properties and efficiency for solving the system in Eq.(1.1). Because of the nice decoupling properties in the IIF method, we will treat the deterministic diffusion and reaction terms in a similar fashion [83], while dealing with the stochastic term explicitly as in the Euler Maruyama method [38]. We compare this approach with similarly constructed schemes whose main difference is in the treatment of the deterministic part of the equation, which can be approximated using ETD, Crank-Nicolson, or Implicit Euler methods. When all of the properties such as order of accuracy, mean errors, and stability region are taken into consideration, the new approach shows many advantages. We also take advantage of the low computational cost of the cIIF methods to similarly construct a stochastic method that can be applied to higher-dimensional problems. The paper is organized as followed. We first present the construction of the method for systems with additive noise or multiplicative noise, along with linear stability analysis and their comparisons with several other methods. Next, we compare the new method with other methods on linear SODEs and SPDEs on their orders of accuracy and stability constraints. Then, we use this approach to study a nonlinear activator-substrate system of two diffusion species and lastly, make our conclusion.

Chapter 2

Ovol2-Zeb1 Loop Governs Multi-step Epithelial-Mesenchymal Transition

2.1 Results

2.1.1 A regulatory network containing Ovol2-Zeb1 mutual repression results in multiple intermediate states and a four-state EMT system

Mutual inhibition loops between EMT-inducing TFs and miRNAs (e.g. Zeb1-miR200 and Snail-miR34a) are critical for robust control of EMT/MET [84]. Recent studies on Ovol2's direct inhibition of Zeb1 [61, 108], coupled with the discovery of the repression of Ovol2's expression by Zeb1 (Appendix A), raise questions about the functions of the Ovol2-Zeb1 mutual inhibition loop in governing the EMT process.

To dissect the role of the Ovol2-Zeb1 loop in EMT dynamics, we incorporated this regulation, as well as the negative regulation of TGFβ signaling by Ovol2 [108] into a framework that

has been successfully used to formulate a 3-state EMT system [115]. The new model thus contains three mutual inhibition loops: Zeb1-miR200, Snail-miR34a and Ovol2-Zeb1, Figure 2.1A. To examine how the system might be stabilized at various stages of EMT, we performed bifurcation analysis with respect to external TGFβ as an EMT inducer. Interestingly, four distinct stable steady states corresponding to four cell phenotypes emerged with the addition of the Ovol2-Zeb1 loop, Figure 2.1B. In particular, two intermediate states appeared between a terminal E state and an M state, Figure 2.1B. We named the intermediate state closer to the E state I_1, and the one closer to the M state I_2. The dynamic feature of the four-state system is consistent with the recently proposed sequential cell-state transition in which more than one intermediate states may exist [99], and it is also compatible with existing EMT models [76, 100, 115] in terms of possible ternary switch in the system (I_1-I_2-M, E-I_1-I_2, E-I_1-M or E-I_2-M, depending on specific external stimulation). Importantly, our model predicted that elevated production of Ovol2 is able to reprogram all other states to the terminal E state, Figure 2.1C. We will return to this notion later.

2.1.2 Stochastic simulations and experimental evidence consistently support the 4-state model and validate the reprogramming ability of Ovol2

From experimental results, Xing Dai's lab independently confirmed the existence of two intermediate states during the transition from epithelial to mesenchymal phenotypes (Further explanations are included in Appendix B and Figure 2.2A). They also concluded that overexpressing Ovol2's gene expression can reprogram I_1 state into an E state as well as reprogram M-state cells into an E state, (Appendix B, Figure 2.2B and E). To compare these observations with our mathematical model, we performed stochastic simulations with the basal model upon fluctuations in gene/protein expression (see details in Methods). We

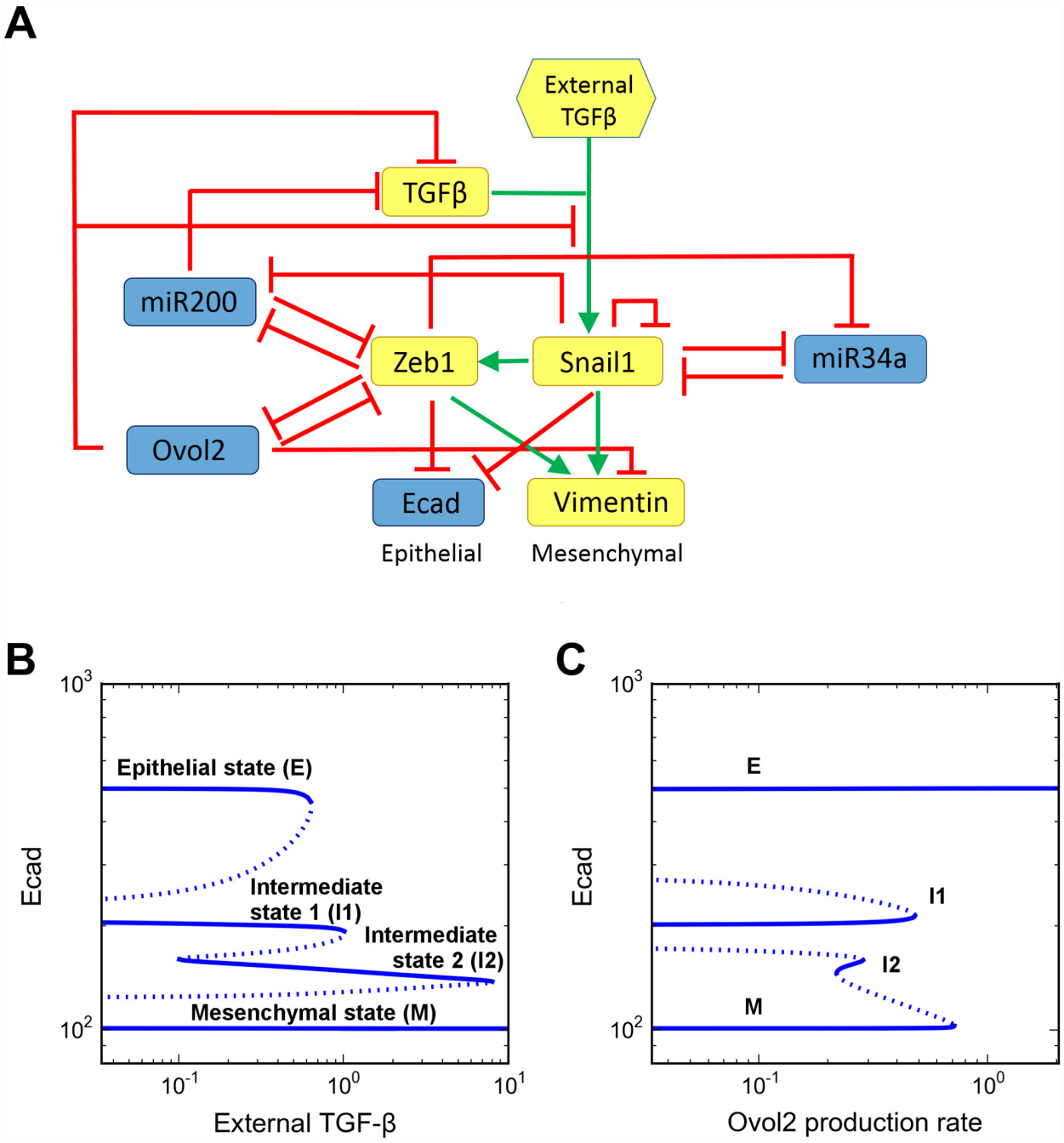

Figure 2.1: Incorporation of an Ovol2-Zeb1 mutual repression module results in the observation of four distinct states in EMT. **A)** Influence diagram of the EMT/MET system. Blue icon: epithelial promoting factor. Yellow icon: mesenchymal promoting factor. Hexagon: extracellular input. **B, C)** One-parameter bifurcation diagrams of E-cadherin (Ecad) with respect to external TGFβ (B) and Ovol2 basal production rate (C). Solid curve: stable steady state. Dashed curve: unstable steady state. In B, only transition between I_1 and I_2 is reversible when TGFβ level is varied. In C, varying Ovol2 alone does not result in any reversible transition, but it can possibly reverse the following TGFβ-induced transitions: E-I_1, E-I_2 and E-M.

started simulation with the initial condition at I_1 state (Figure 2.2C, green). Remarkably, at high basal production rates of Zeb1 and Ovol2 (Figure 2.2C, red and blue), the simulation produced similar Ecad/Vim expression patterns of these populations as those observed in the experiments. These observations demonstrate the bidirectional differentiation potential of MCF10A cells towards two opposite directions (i.e., I_1 to E or I_1 to M) and provide evidence for the opposing roles of Ovol2 and Snail/Zeb1 in the dynamic EMT system of MCF10A cells. In addition, results of stochastic simulations for control M-state cells and Ovol2-overexpressed M-state cells using the corresponding conditions are in good agreement with experimental results, Figure 2.2F. Interestingly, a time series experiment revealed that the downregulation of Vim precedes the induction of Ecad, suggesting that upon Ovol2 expression these cells first lose their memory of the M state and then acquire the E phenotype, Figure H.4.

Consistent with previous findings [115], a high dose of TGFβ resulted in a complete conversion of cells to what appears to be the M state, whereas low dose of TGFβ induced the appearance of two new populations in a heterogeneous culture that are likely I_2 (previously termed P state for partial EMT in Zhang et al. [115]) and M states (Figure 2.3A and B and Figure 2.4). Of note, in both mathematical modeling and experiments, the I_2 state appears less stable than I_1, as it 1) shows more vulnerability when facing fluctuations (Figure 2.4); 2) entails a narrow range of Ovol2 concentration in the absence of strong TGFβ signaling (Figure 2.1C); and 3) is barely distinguishable from the M or I_1 state experimentally and in simulations (Figure 2.3A and Figure 2.4). The degree of separation of the different cell populations in our study is less remarkable than that reported [115], possibly due to the dynamic nature of I_2 and the subtle differences in experimental conditions.

Taken together, these experimental results support our computational discovery of a four-state dynamic system. Moreover, they highlight the ability of Ovol2 in reprogramming both I_1-and M-state cells into a terminal E state, as predicted by our model. The EMT phenotypes and the cell state transitions that we have discovered through modeling and experiments are

summarized in Figure 2.3C.

2.1.3 Critical regulatory controls in maintaining the four states

First, we explored the roles of Ovol2 in regulatory control of the four states. Through bifurcation analysis with respect to external TGFβ and basal production rate of Ovol2 representing examples of EMT-inducing and -suppressing signals that are responsive to changes of the tissue microenvironment, we found these two signals to produce various combinations of cell phenotypes (Figure 2.5A(I)). Clearly, Ovol2 basal production rate exerted positive and negative effects on the robustness of E and M states, respectively (Figure 2.5A, blue and pink areas), and this is consistent with the demonstrated role of Ovol2 in preventing EMT and inducing MET ([91, 108] and this study). The effect of Ovol2 on the two intermediate states can be either positive or negative. Stability of I_1 can be maintained when the strengths of Ovol2 and TGFβ signals are approximately proportional, with low levels of both giving rise to the most robust condition (Figure 2.5A, cyan area). In contrast, stability of I_2 requires a minimum rate of Ovol2 basal production, but its robustness increases with higher levels of both Ovol2 and TGFβ (Figure 2.5A, orange area). In a specific case, when TGFβ signal was increased by 10-fold, higher Ovol2 basal production rate was required to retain the stability of both I_1 and I_2 states and to prevent the cells from differentiating into M state (Figure 2.5B, blue and orange triangles). Conversely, when Ovol2 basal production rate was increased, higher TGFβ signal strength was required to retain two stable intermediate states and to prevent the cells from differentiating into E state (Figure 2.5C, blue and orange diamonds). Overall, our analysis suggests that Ovol2 production tends to stabilize E state and destabilize M state, and that the two intermediate states are favored in two distinct conditions (high- versus low-signals) but both require the proper balance between EMT-inducing (e.g. TGFβ) and -suppressing signals (e.g. signals that induce Ovol2 expression).

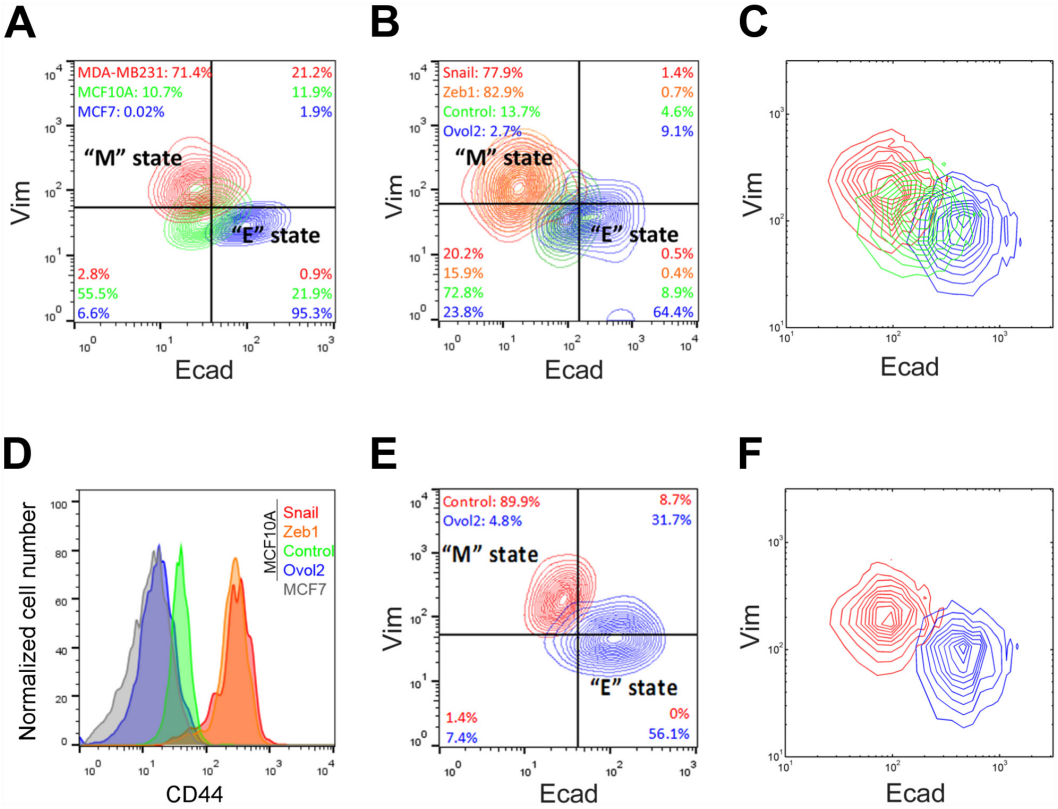

Figure 2.2: Experimental evidence for bidirectional potential of MCF10A cells. **A, B, D, E)** Flow cytometric analysis of epithelial marker (Ecad) and mesenchymal marker (Vim) profiles. **A)** Direct comparison of MCF10A with luminal (epithelial)-type cancer cell line MCF7 and basal (mesenchymal)-type cancer cell line MDA-MB231. MCF10A (green) falls in the intermediate state between MCF7 (blue) and MDA-MB231 (red). Analyses were performed at 90 to 100% confluency. **B)** Bidirectional potential of MCF10A cells. E(I)MT and M(I)ET was induced by forced expression of transcription factors Snail or Zeb1, and Ovol2, respectively. After 6 days of lentiviral infection, Snail (red) and Zeb1 (orange) induced EMT while Ovol2 (blue) induced MET as compared to the empty vector control (green). **C)** Stochastic simulations for a population of 2000 cells in three different conditions: basal parameter set (green), high basal production rate of Zeb1 *ZEB1* (red, *Zeb1* mRNA basal production rate was raised to 0.01 μM/hr) and high basal production rate of Ovol2 (blue, Ovol2 basal production rate was raised to 2 μM/hr). Initial conditions are all at I_1 state. **D)** The histogram shows the expression status of CD44. M states (Snail; red and Zeb1; orange) correlate with high CD44 expression while cells in E state (Ovol2; blue) show low CD44 expression as compared to empty vector control (green). MCF7 is shown as a representative cell type in the E state with low CD44 expression. **E)** Ovol2 reprograms MDA-MB231 cells from M- to E- state. Cells were analyzed after 6 days of control (red) or Ovol2-expressing (blue) lentiviral infection. **F)** Stochastic simulations with a basal parameter set (red) and high basal production rate of Ovol2 (blue, Ovol2 basal production rate was raised to 2 μM/hr). Initial conditions are all at M state.

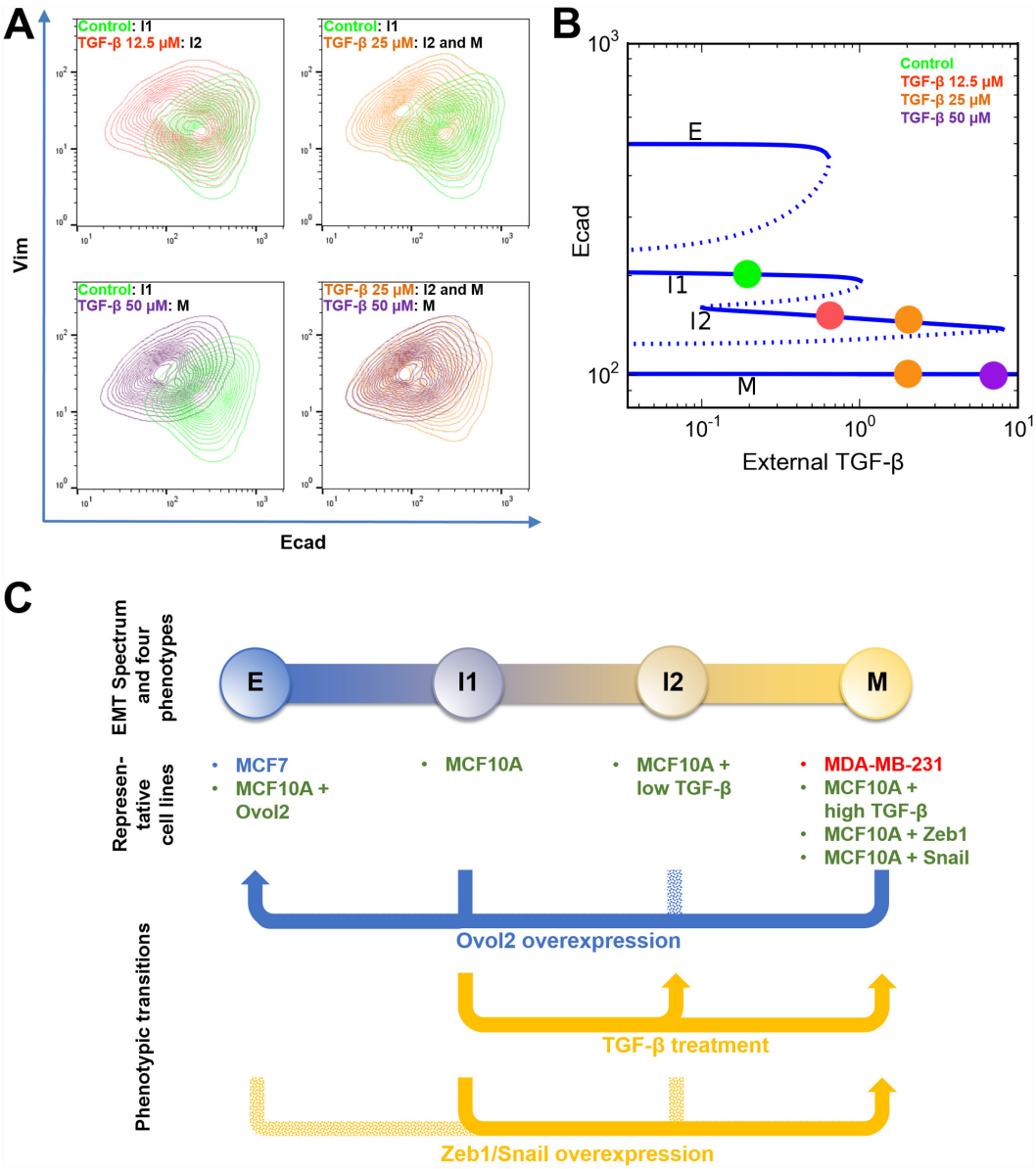

Figure 2.3: Stepwise induction of EMT in MCF10A cells by different doses of TGFβ. **A)** Cells were treated with various concentrations of TGFβ for 10 days and analyzed for Ecad/Vim expression by flow cytometry. Each panel shows a superimposed image of two treatment conditions. Note that the 25 μM TGFβ treatment gave rise to a heterogeneous population containing I_2 cells and M cells (orange). **B)** The corresponding steady states of the cellular phenotypes (indicated as colored dots) observed under various TGFβ concentrations in A are mapped to the bifurcation diagram shown in Figure 2.1B. **C)** An illustrative summary of phenotypic transitions in the four-state system. Solid arrow represents transition with experimental evidence from this study. Dotted arrow represents hypothetical transition without experimental evidence.

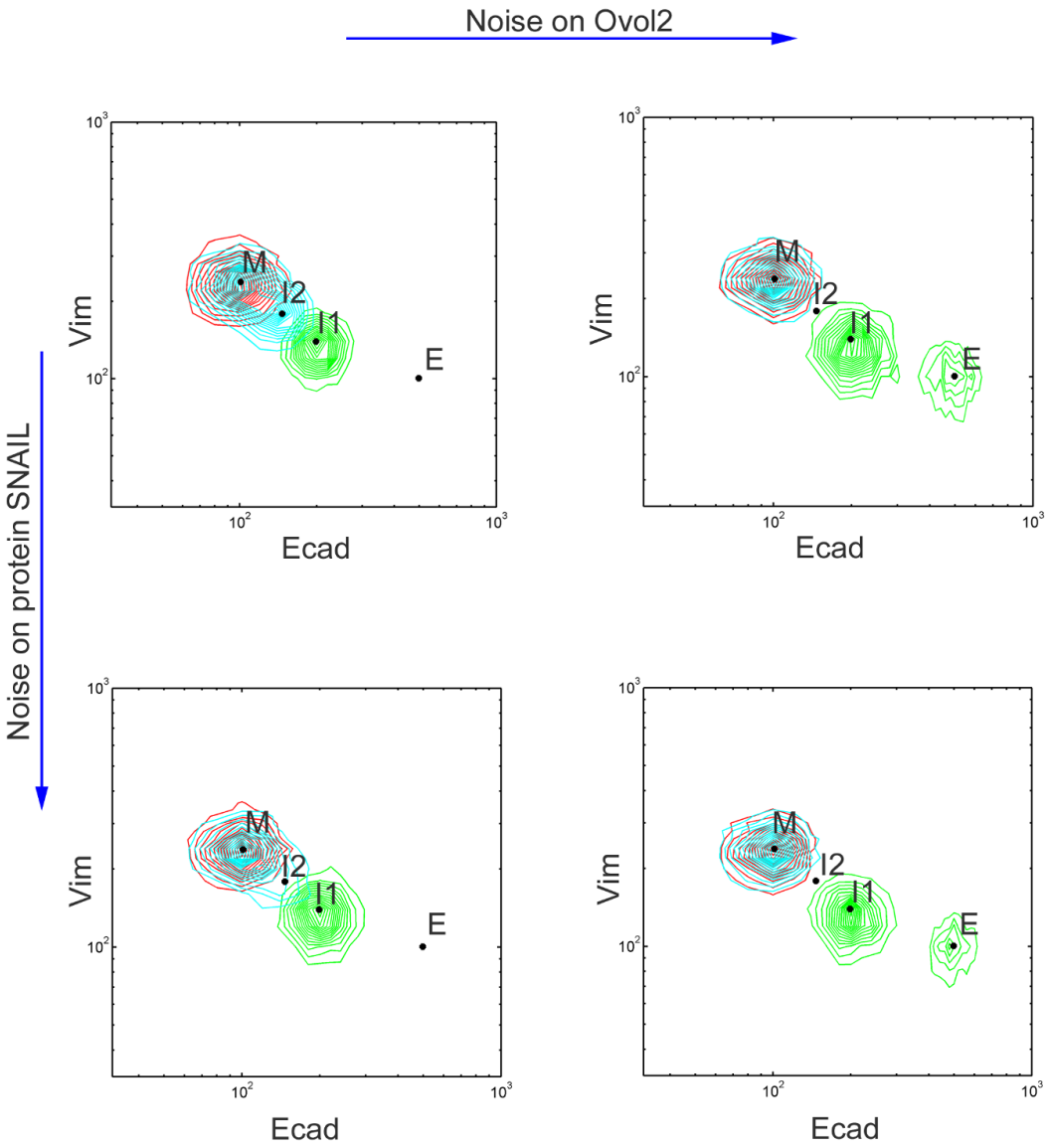

Figure 2.4: Stochastic simulations for stepwise I_1-I_2-M transitions upon TGFβ treatment. Stochastic simulation (started at I_1 state) for a population of 2000 cells at three concentration of TGFβ. Green: no TGFβ (I_1 state). Red: high (10 units) TGFβ concentration (M state). Cyan: intermediate (2.5 units) TGFβ concentration (a mixture of I_2 and M populations can be obtained at the low-noise condition). At high TGFβ concentration, the system is monostable at M state. At intermediate TGFβ concentration, the system is bistable at M or I_2 state. See Figure 2.1B.

Next we reduced the strength of the Zeb1-Ovol2 mutual inhibition loop to determine its specific role in the four-state system. This led to no significant effect on the robustness of E state, moderately positive effect on that of M state (Figure 2.6, middle column, blue and pink areas), but significantly reduced robustness of the two intermediate states (Figure 2.6, middle column, cyan and orange areas). A complete blockage of the mutual inhibition loop resulted in a very small I_1 region, and complete disappearance of the I_2 region (Figure 2.6, right column). The role of the Ovol2-Zeb1 loop appeared distinct from that of the miR34a-Snail and miR200-Zeb1 mutual inhibition loops, as at least one of the miRNA-TF loops becomes dispensable for I_2 (but not I_1) when the other is intact (Figure 2.7, left and middle columns). This said, a complete blockage of both miRNA-TF loops abolished the intermediate states (Figure 2.7, right column). Partial blockage of each mutual inhibition loop gave rise to complex effects on the robustness of the two intermediate states (Table 2.1 and Figure H.6), but these effects are consistent with the finding of redundancy between the two miRNA-TF loops in terms of maintaining I_2 state. Interestingly, partial blockage of miR200-Zeb1 resulted in a merge of the I_1 and I_2 regions, forming a large, continuous intermediate region (Figure H.6). Collectively, these results suggest that while all three mutual inhibition loops contribute to the existence and robustness of the two intermediate states, the strength of the Ovol2-Zeb1 loop is more critical.

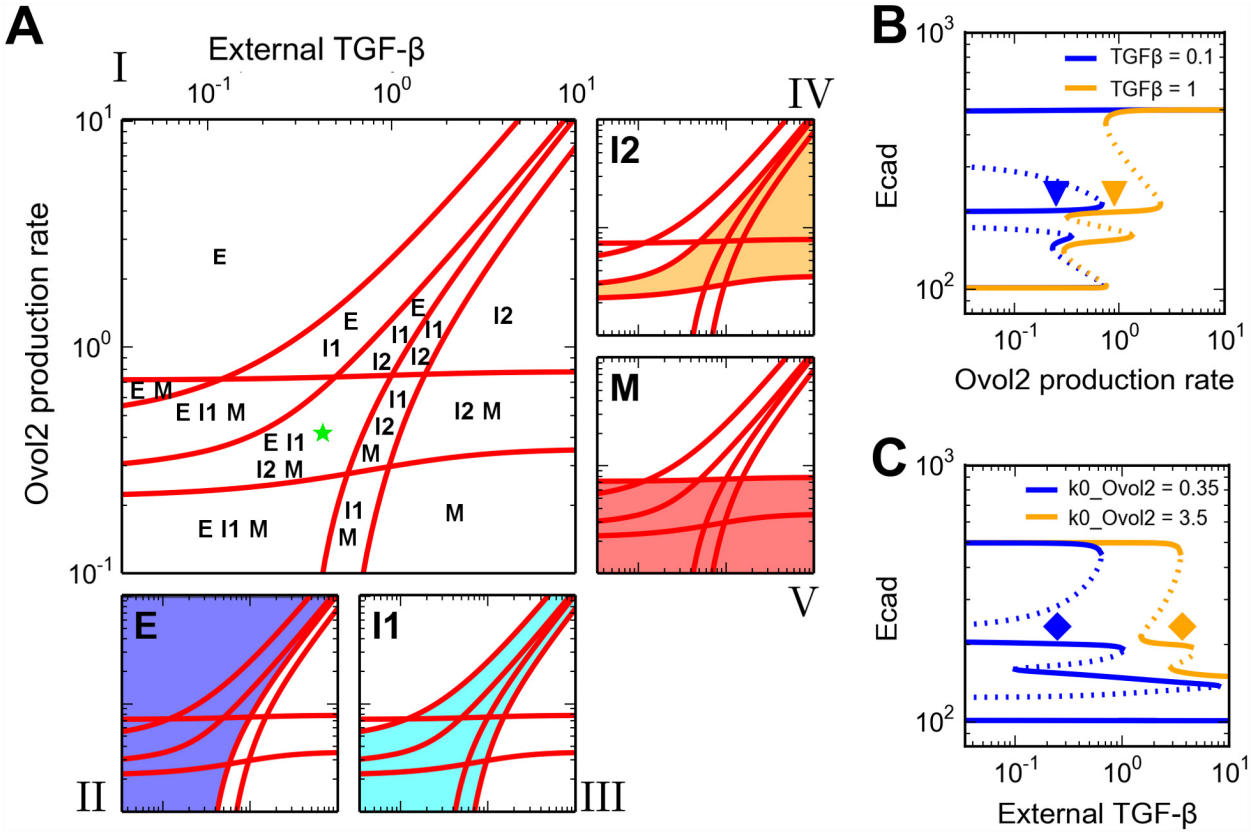

Figure 2.5: Roles of EMT-promoting and -inhibiting factors in the four-state EMT system. **A)** Two-parameter bifurcation diagram with respect to external TGFβ and Ovol2 basal production rate. The red curves were computed by extending the saddle-node bifurcation points obtained in one parameter bifurcation analysis (Figure H.5), and they define different parameter regions that can be mono-stable, bi-stable, tri-stable or tetra-stable depending on the number of possible stable phenotypes (see labels), and each multi-stable region can be viewed as an area where multiple phenotypes co-exist (II-V). The size of each phenotype region is an indication of robustness of the phenotype when the two signals are varied. Green star: a basal parameter set and an intermediate TGFβ concentration that together give rise to four phenotypes. **B, C)** One-parameter bifurcation diagrams of Ecad with respect to external TGFβ (B) and Ovol2 basal production rate (C). Solid curve: stable steady state. Dashed curve: unstable steady state. A basal parameter set (blue) and a perturbed parameter set (orange) are compared in each plot. Triangles and diamonds denote the conditions under which both I_1 and I_2 are stable.

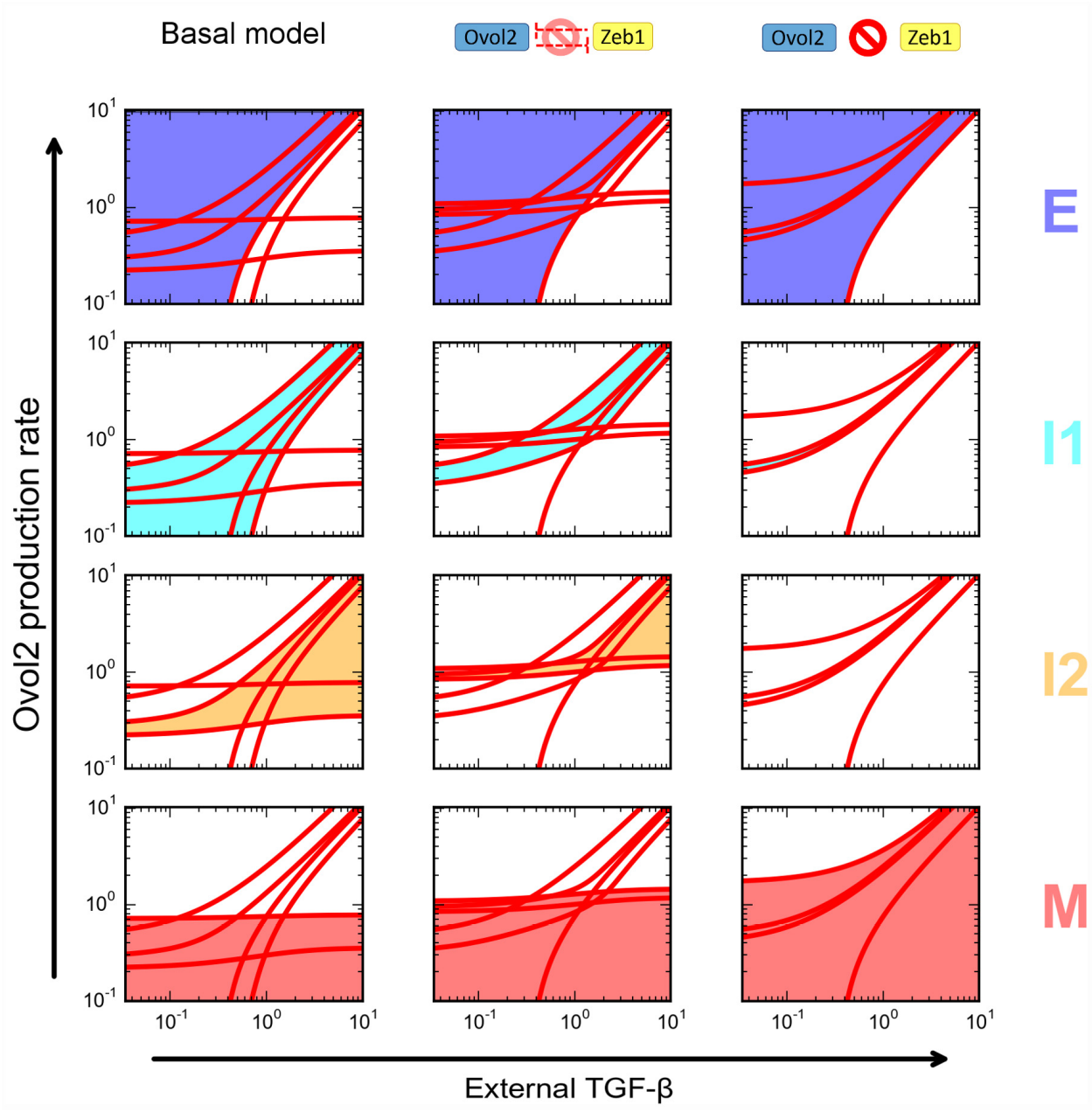

Figure 2.6: Roles of the Ovol2-Zeb1 mutual inhibition loop in the four-state EMT system. Comparison of the basal model (left column), reduced Ovol2-Zeb1 mutual inhibition (middle column), and blocked Ovol2-Zeb1 mutual inhibition (right column) on the four phenotypes. Each subplot is a two-parameter bifurcation diagram similar to Figure 2.5A. Subplots in each column highlight the various phenotypes in one condition. Shaded areas are highlighted phenotypes. Colors of the shading correspond to the colored labels on the right.

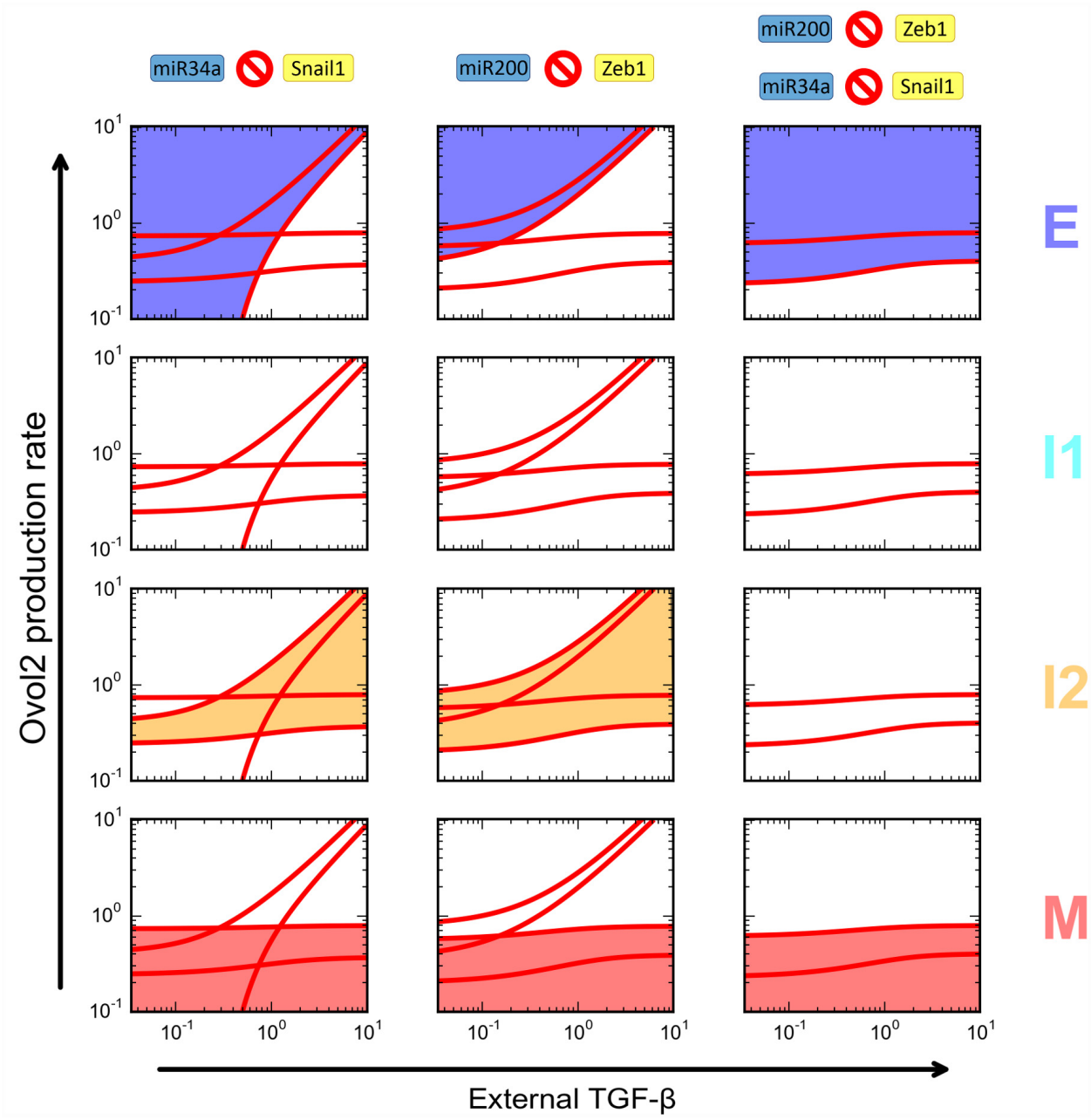

Figure 2.7: Roles of the miR34a-Snail and miR200-Zeb1 mutual inhibition loops in the four-state EMT system. Comparison of removing miR34a-Snail mutual inhibition (left column), miR200-Zeb1 mutual inhibition (middle column), or both (right column) on the four phenotypes. Each subplot is a two-parameter bifurcation diagram similar to Figure 2.5A. Subplots in each column highlight the various phenotypes in one condition. Shaded areas are highlighted phenotypes. Colors of the shading correspond to the colored labels on the right.

2.1.4 Distinct differentiation propensities of the two intermediate states

We next used stochastic modeling to examine how likely a population of cells at the two intermediate states differentiates into E or M state when gene/protein expression fluctuates. We chose a condition under which the four states are stable (Figure 2.5A, green star). Simulations were performed under this condition for two populations of cells originating from I_1 and I_2 respectively (Figure 2.8A and F). When fluctuations were small, cells stayed in the basins of attraction of their initial steady state by the end of the simulation (Figure 2.8A and D). Large fluctuations triggered both E(I)MT and M(I)ET of the cells originally in I_1, resulting in a heterogeneous population containing E and M phenotypes, whereas the same level of fluctuations triggered E(I)MT of the cells originally in I_2 (Figure 2.8B and E and Figure H.7). Thus, the I_1 and I_2 cells have distinct differentiation propensities, with I_1 cells more likely to differentiate into E state, whereas I_2 cells more likely into M state. These simulations also allowed us to infer that E and M states are more stable than the intermediate states, and that I_1 is more stable than I_2 under the particular conditions tested (Figure 2.8C and F).

We also asked whether changing Ovol2 production rate can affect the differentiation propensities from the I_1 state. We reduced the basal production rate of Ovol2 by 20% and performed stochastic simulations starting from I_1 as in Figure 2.8A and B. We found that this reduced Ovol2 production rate enabled more cells to settle at M state, and fewer cells to settle at I_1 or E state (compare Figure 2.8B and H). We speculate that this is due to the reduced stability of E and I_1 states, and/or the reduced energy barrier from I_1 to I_2 and M states (Figure 2.8I), providing a possible thermodynamic explanation for the role of Ovol2 in preventing EMT and inducing MET. Conversely, increased basal production of Ovol2 enabled some of the cells from I_2 state to settle at E state instead of M state in the presence of large fluctuations (Figure 2.8J, K and L; compare panels E and K). These results suggest that

Mutual inhibition loop	*Knockdown condition*	*Example of parameter setting*	*Major effect on intermediate states*
miR34a-Snail	Partial removal	$0.1\%1/J2_{200}$	Disappearance of I_1 and decrease of I_2
		$25\%K_{SR}$, $25\%1/J1_{34}$	Decrease of both I_1 and I_2
		$50\%K_{SR}$, $50\%1/J1_{34}$	Decrease of I_1
		$0.1\%K_{SR}$	Disappearance of I_1 and increase of I_2
	Complete removal	$0.1\%K_{SR}$, $0.1\%1/J1_{34}$	
miR200-Zeb1	Partial removal	$90\%1/J2_{200}$	Decrease of I_2
		$75\%1/J2_{200}$	Merge of I_1 and I_2
		$50\%K_1$	Decrease of I_1 and increase of I_2
		$25\%K_1$, $25\%1/J2_{200}$	Disappearance of I_1 and increase of I_2 i.e *I_1 and I_2 merge to form a large intermediate state where this intermediate state is closer to I_2 than I_1*
	Complete removal	$0.1\%K_1$, $0.1\%1/J2_{200}$	
miR34a-Snail and miR200-Zeb1	Complete removal	$0.1\%K_{SR}$, $0.1\%1/J1_{34}$, $0.1\%K_1$, $0.1\%1/J2_{200}$	Disappearance of both I_1 and I_2
Ovol2-Zeb1	Partial removal	$50\%1/J_O$	Increase of I_2
		$50\%1/J2_{zeb}$	Decrease of I_1 and I_2
		$25\%1/J_O$, $25\%1/J2_{zeb}$	Decrease of I_1 and disappearance of I_2
	Complete removal	$0.1\%1/J_O$, $0.1\%1/J2_{zeb}$	Tiny I_1 region and disappearance of I_2

Table 2.1: Influence of blocking mutual inhibition loops on the two intermediate states

the differentiation propensities of the two intermediate states can be regulated by tuning the level of Ovol2 expression.

2.2 Discussion

Our study provides both modeling and experimental evidence for a new intermediate state that lies between E and M states in addition to the recently observed intermediate state [76, 100, 115]. Previous studies based on epigenetic modifications predicted that multiple intermediate states may exist between terminal E and M states, and they may contribute

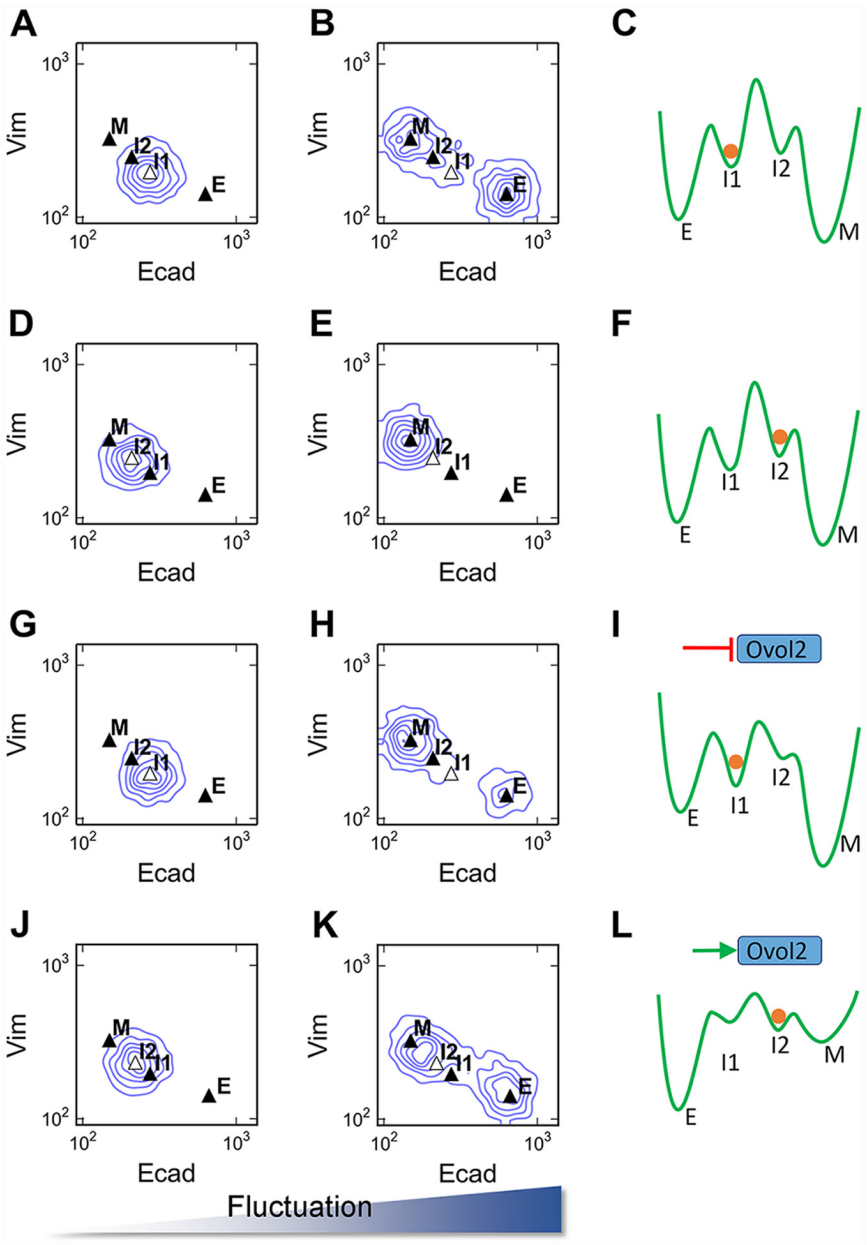

Figure 2.8: Distinct differentiation propensities of the two intermediate states. **A, B, D, E)** Stochastic simulations for a population of 5000 cells in four different conditions. Basal parameter set and intermediate external TGFβ concentration (0.5) were used (green star in Fig 6A). **A)** Initial condition: I_1; small fluctuations. **B)** Initial condition: I_1; large fluctuations. **D)** Initial condition: I_2; small fluctuations. **E)** Initial condition: I_2; large fluctuations. **G, H)** Stochastic simulations for a population of 5000 cells initially at I_1 state. Ovol2 basal production level was reduced by 20% from basal parameter. **J, K)** Stochastic simulations for 5000 cells initially at I_2 state. Ovol2 basal production level was increased by 100% from basal parameter. **C, F, I, L)** Metaphoric energy landscapes (green curves) for I_1 **(C)**, I_2 **(F)** initial conditions, and reduced **(I)** or increased **(L)** Ovol2 basal expression rate. Orange circle represents the initial condition.

to phenotypic plasticity in a continuous manner [99]. Additionally, Huang et al. classified 43 ovarian carcinoma cell lines into four subgroups, including an E-like intermediate and an M-like intermediate states in the EMT spectrum, based on expression patterns of signature EMT genes [42]. To our knowledge, our work is the first unequivocal demonstration of two intermediate states in EMT. Previous theoretical study revealed four types of stable states during T cell differentiation [41]; a common feature of that and our study is the inclusion of multiple (a minimum of three) positive feedback loops (including mutual inhibition).We anticipate that as the complexity of modeling increases by adding more regulatory elements, even more intermediate states may be observed, with the most extreme scenario being a spectrum of metastable or stable cell phenotypes lying between the terminal E and M states. A unique and interesting feature of our four-state model is that intermediate states are not necessarily metastable; instead they can be stable with no (I_1) or high (I_2) EMT-inducing/inhibitory forces. It is the balance between these two opposing forces that is critical for maintaining the intermediate states.

Given the assumptions we make in our mathematical model, we have shown that the Ovol2-Zeb1 mutual inhibition loop is necessary for maintaining a four-state system. On one hand, this highlights the unique importance of the role of Ovol2 in EMT. On the other hand, the model leaves open the possibility that a four-state system could be governed by other unknown TFs that might be involved in a similar mutual inhibition loop. As discussed above, with additional positive feedback loops (> 3), it is conceivable that additional intermediate states will emerge. As such, our model provides a framework for identifying and analyzing multiple intermediate phenotypes in EMT, and suggests a general and unique role of TF-based mutual inhibition loop in this system. With a proposed Ovol-Zeb1 mutual inhibition loop, a recent modeling study suggested important roles of Ovol2 in controlling the previously established three-state EMT system [46]. This is in agreement with our findings that Ovol2 is critical for both intermediate states.

What is the advantage of having intermediate state(s)? Such state can be a "hybrid" state,

where cells exhibit both, albeit partial, "E" and "M" phenotypes. Indeed, during mammary gland morphogenesis, some epithelial cells at the tips of growing ducts express mesenchymal markers while simultaneously retaining epithelial integrity, suggesting that they are in a naturally occurring "hybrid" state [27]. In this case, a "hybrid" phenotype would enable the cells to undergo collective migration, by which they invade the surrounding stroma as a coherent epithelial front to facilitate branching morphogenesis. The same may be true for metastatic cancer cells as they acquire a mesenchymal phenotype to invade the surrounding tissue and colonize distant sites as epithelial tumors. Alternatively, an intermediate state can be a "naïve" state, where cells are devoid of typical epithelial and mesenchymal features. Along this line, we note our experimental observation that MCF10A cells seem to first lose their initial phenotypes (E or M), and then gain their destination phenotypes (M or E) during the factors-directed state transitions (Figure 2.2). Traveling through a "naïve" state could be a useful mechanism to erase memories of old lineages, thus creating a window of opportunity for expanded differentiation plasticity as desired for multipotent stem cells.

Why multiple intermediate states then? Chuong and Widelitz proposed the interesting idea that stem cell states can be regulated depending on the physiological needs of tissues to generate different numbers of intermediate stops on their journey to differentiation [16]. The same concept may be applicable to the EMT system, as having multiple intermediate states offers additional facets of regulation to more precisely control the temporal and spatial flux of epithelial cells through their differentiation/dedifferentiation pathway to adapt to various tissue environments or topology. Regardless of whether cells adopt an intermediate fate to gain "hybrid" behavior (e.g., collective migration) or to dedifferentiate to a naive state for lineage plasticity, the more intermediate states there are, the more thermodynamic traps that would be. Thus having more than one intermediate state would create a more controllable energy barrier so that cells do not easily fall into the mesenchymal state, which we know from previous studies is largely irreversible [115]. Non-genetic heterogeneity and spontaneous conversion among subpopulations have been documented for hematopoietic stem

cells and breast cancer cells [12, 13, 32, 67]. Theoretical analysis of these dynamic processes often involves the assumption that gene regulatory networks can generate multiple stem-like states that are adjacent in state space [13, 32, 43, 119]. Our model presents a good example in which a network of three mutual inhibition loops is capable of giving rise to two adjacent states that may be associated with subpopulations of cells having distinct propensities for differentiation or tumorigenesis. The unstable nature of the I_2 state under conditions examined and the phenotypic similarities between I_1 and I_2 states prevented us from further characterizing the molecular differences between the two intermediate cell populations and their corresponding cellular behaviors. As such, the functional significance of having two intermediate states has yet to be experimentally established.

We have demonstrated bidirectional transitions of MCF10A cells upon Zeb1/Snail overexpression (I_1-M transition), TGFβ treatment [I_1-(I_2)-M transition] and Ovol2 overexpression (I_1-E transition) (Figure 2.3B). It is tempting to ask which extracellular signaling molecules can trigger Ovol2 upregulation and the subsequent transition to E state under physiological conditions. Among the possible candidates is a BMP signal as it is known to induce MET [91, 108] and to positively regulate Ovol2 expression during embryonic stem cell differentiation [117]. Identification of Ovol2-inducing external signals that can induce MET of MCF10A cells will enable a finer analysis of the dynamic process of MET as well as further experimental validation of our mathematical model.

In summary, our work identifies transcriptional factor Ovol2 and its mutual inhibition relationship with Zeb1 as critical additions to the known EMT regulatory network. Specifically, these new regulatory elements are important for attaining and maintaining the two intermediate states. Furthermore, their experimental perturbations allowed us to observe the bidirectionality of transitions from the intermediate states. Together, our study offers a framework for understanding the molecular strategies and design principles by which epithelial stem, progenitor, or cancer cells achieve multipotency or collective migration.

Chapter 3

Multiple Cell States Reduce Noise in EMT

3.1 Mathematical models and stochastic simulation of multiple-state EMT models

3.1.1 Model construction

We explored the functional significance of the intermediate states in the EMT process by adapting the population dynamics to modelling the transition from epithelial to mesenchymal phenotypes. Each steady state in the transition was portrayed by a population of cells that assumed the phenotype reflective of that state. In other words, an n-state EMT system has n different populations of cell, with one of them being epithelial cells and another being mesenchymal cells. Hence there is a total of $n-2$ groups that represent the cell populations of different intermediate stages. In this paper, we discussed the population dynamics for four different systems: two-state, three-state, four-state, and five-state EMT processes (Figure

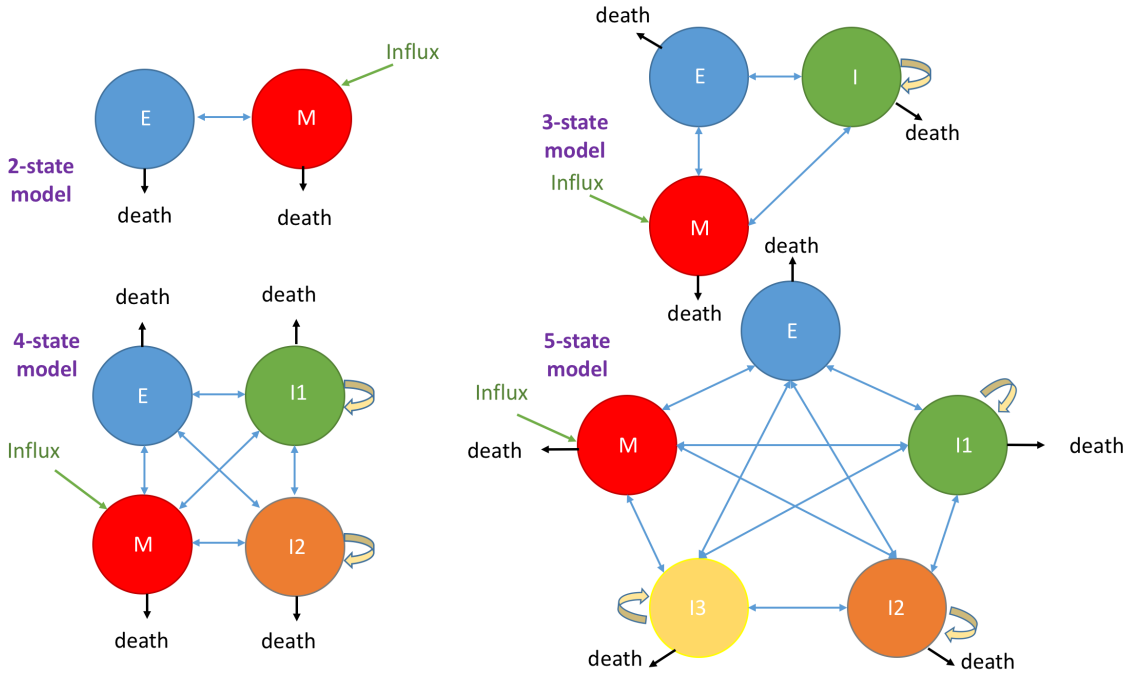

Figure 3.1: Four population dynamics models for four distinct EMT systems: two-state, three-state, four-state, and five-state transitions. In each model, E denotes the epithelial population, M mesenchymal population, and I_i the population corresponding to the i^{th} intermediate state. In each model, the blue arrows describe the transitions of cells between two populations, where the arrow heads describe the direction of each transition. The black arrows indicate cell death, the yellow arrows the self-renewal of (stem) cells in the intermediate states, and the green arrows that point towards the mesenchymal population indicate a constant influx of cells into this population.

3.1). In each system, each population of cells was characterized by death rate and transition rates from/to another population. The populations that corresponded to the intermediate states were given stem cell features, described by self-renewal rates and lower death-rates than other types of cells. We modeled the population of each cell phenotype with a linear ordinary differential equation (ODE), where we introduced a constant influx of cells to the mesenchymal population under the assumption that mesenchymal cells are mobile and invasive. With this influx, all of the matrix systems in our models are nonsingular. The ODE that depicts the change in the population size of a particular cell phenotype P_i, $i = 1, 2, ..., n$, is as follows:

$$\frac{dP_i}{dt} = \sum_{k \neq i} \alpha_{ki} P_k - \sum_{k \neq i} \alpha_{ik} P_i - d_i P_i + s_i P_i \tag{3.1}$$

where α_{ki} is the cellular conversion rate from P_k to P_i, d_i is the death rate, and s_i is the self-renewal rate where $s_i = 0$ if the phenotype of the cell population P_i is epithelial or mesenchymal. The cellular conversion rates, death rates, and self-renewal rates all assume constant values. In addition, if P_i describes the population of mesenchymal cell, then

$$\frac{dP_i}{dt} = \sum_{k \neq i} \alpha_{ki} P_k - \sum_{k \neq i} \alpha_{ik} P_i - d_i P_i + s_i P_i + I_M \tag{3.2}$$

with I_M being the constant influx of cells into the mesenchymal population. For our simulations, we chose the range for our parameter values to be between 1e$-$5 and 10, with the death rates d_i of the "stem cell" populations to be between 1e$-$5 and 1e$-$2 to reflect the lower death rates of stem cells. Meanwhile, since the epithelial and mesenchymal populations were assumed to have higher death rates, we chose the death rates to be between 0.05 and 10. The quantity of cell population is in arbitrary unit. One time unit in our model corresponds to one day.

In order to explore the noise attenuation properties of each population dynamics model, we introduced multiplicative noise $\sigma_i P_i dW_i$ to our ODE system via multiple means, where σ_i is

the noise coefficient and W_i describes the Standard Brownian motion. This noise is referred to as "input noise" in this paper. First, we added noise to the epithelial and mesenchymal populations, then observed the effects that noise introduction had on each individual population as well as overall. Next, we added noise to one intermediate population besides the two existing noises on epithelial and mesenchymal cells, and finally we introduced multiplicative noise to each cell population. To ensure consistency between the population fluctuations, we chose the noise coefficient $\sigma_i = 1$ for all $i = 1, 2, ...n$.

3.1.2 Model comparison through Differential Evolution algorithm

To compare various models with distinct number of parameters, we used Differential Evolution (DE) to optimize these models with respect to their ability to attenuate noise, and we subsequently compared the optimized parameter sets in terms of the noise attenuation property. DE is a method for improving a set of parameter values (i.e., a parameter vector $\mathbf{x}$, as above) with respect to an objective function, $\Phi(\mathbf{x})$, by generating a sequence of trial parameter vectors by processes of reproduction and selection [89]. Reproduction generates an offspring parameter vector $\mathbf{u}$ from a parent parameter vector $\mathbf{x}$ by diversification. If the offspring performs better than the parent, then the cost value produced by the objective function for the offspring is less than that of the parent, i.e. $\Phi(\mathbf{u}) < \Phi(\mathbf{x})$. In this case, the parent vector $\mathbf{x}$ is replaced by the offspring vector $\mathbf{u}$ in the next generation. The DE algorithm was previously shown to be efficient in optimizing models for biological systems [39]. With this algorithm, we iteratively searched for a newer set of parameter values that provided a better cost to the objective function $\Phi(x)$ than the previous set until the cost converged to the most optimal value. We chose the cost to be the average coefficient of variation of the population size for all phenotypes, i.e. the sum of the coefficients of variation of all the cellular populations divided by the number of states in a particular EMT system

that we study. As a result, a lower cost value implies an overall reduction in fluctuations in each cell population and is thus desirable. We applied the optimization algorithm ten times to obtain ten sets of parameters with their corresponding cost values.

During DE, parameter vectors evolve from generation to generation. Each generation (indexed by $t = 0, 1, ...$) consists of N parameter vectors $\mathbf{x}_j$, $j = 1, ..., N$. Hence, the real number $x_{ij}(t)$ is the value of the i^{th} parameter in the j^{th} parent in the t^{th} generation. Let $\mathbf{u}_j(t)$ be the parameter vector for the single offspring of the j^{th} parent in the t^{th} generation. The components of this vector, $u_{ij}(t)$ for $i = 1, ..., D$, are constructed in two steps (mutation and crossover). Then, given the two parameter vectors $\mathbf{x}_j(t)$ and $\mathbf{u}_j(t)$, a decision is made as to which one is propagated to generation $t + 1$. The following three operations propagate parameter vectors from one generation to the next:

1. Mutation: First, we create a "mutant" vector $\mathbf{v}_j(t)$ by perturbing a parental vector $\mathbf{x}_j(t)$. In our DE approach, we let the perturbation vector be the difference between the parameter vectors of two additional parents, j' and j'', chosen at random from the t^{th} generation of parents. All three parents must be different. The "mutant" vector is defined by:

$$\mathbf{v}_j(t) = \mathbf{x}_j(t) + F \cdot (\mathbf{x}'_j(t) - \mathbf{x}''_j(t)) \tag{3.3}$$

   where F is a scalar, $0 < F < 1$, that determines how "aggressive" the mutation is.

2. Crossover: Next we allow for crossover between the parental parameter set $\mathbf{x}_j(t)$ and the mutant parameter set $\mathbf{v}_j(t)$. Component-wise, the offspring vector $\mathbf{u}_j(t)$ receives a parameter value from the mutant vector with probability C (the crossover probability)

or from the parent vector with probability $1 - C$:

$$u_{i,j}(t) = \begin{cases} v_{i,j}(t), \text{ if } rand(0,1) \leq C \\ x_{i,j}(t), \text{ otherwise} \end{cases} \tag{3.4}$$

$i = 1, 2, ..., D$ and $j = 1, 2, ..., N$.

We choose $C = 0.5$ so that parental values and mutant values have equal chances to be in the offspring vector.

3. Selection: Depending on their relative performances, $\mathbf{x}_j(t)$ or $\mathbf{u}_j(t)$ passes on to the next generation. There are two possibilities here. For "greedy" selection, the offspring replaces its parent if it is superior:

$$\mathbf{x}_j(t+1) = \begin{cases} \mathbf{u}_j(t), \text{ if } \Phi(\mathbf{u}_j(t)) < \Phi(\mathbf{x}_j(t)) \\ \mathbf{x}_j(t), \text{ otherwise} \end{cases} \tag{3.5}$$

For "non-greedy" selection, the condition for replacement is:

$$\Phi(\mathbf{u}_j(t)) \leq \Phi(\mathbf{x}_j(t)) \tag{3.6}$$

We implement DE using the Matlab code available in Price and Storn's algorithm for DE [95].

3.1.3 Application of Fluctuation Dissipation Theorem

To measure the fluctuations on each cell population, we calculated the coefficient of variation of the population size for each cell phenotype from the covariance matrix that we found using the Fluctuation Dissipation Theorem [52, 58]. The Fluctuation Dissipation Theorem helps to predict the behaviors of a system in thermal equilibrium by establishing a relation between the frictional force and random force created by the Brownian particle's impacts with surrounding molecules. We applied the version of the Fluctuation Dissipation Theorem mentioned in [51] to our system of linear stochastic ODEs:

$$d\mathbf{P} = A\mathbf{P}(t)dt + \Gamma(\mathbf{P}(t))d\mathbf{W_t} \tag{3.7}$$

where A is a constant matrix and $\Gamma(\mathbf{P}(t))d\mathbf{W_t}$ represents the multiplicative noise term. $\Gamma(\mathbf{P}(t))$ is a diagonal matrix whose i^{th} entry is $\sigma_i P_i(t)$. With this application of the Fluctuation Dissipation Theorem, we found the covariance matrix Σ for the equilibrium system through the formula

$$A\Sigma + \Sigma A^T = -\Gamma^2(\mathbf{P_{ss}}(t)) \tag{3.8}$$

with $\mathbf{P_{ss}}(t)$ being the steady states of the deterministic system $d\mathbf{P}/dt = A\mathbf{P}(t)$.
To quantify the noise attenuation performance of each EMT system, we aggregate the results of ten different optimization simulations using the DE algorithm [89] and calculate the mean and standard deviation of the coefficients of variation that describe the fluctuations in a specific population. Figures 3.2-3.8.

3.1.4 Sensitivity analysis

In our sensitivity analysis, we defined our parameters as follows: k_{ij} describes the cellular transition rate from the population in the i-state to the population in the j-state where i, j = E, M, I, 1, 2, and 3. Here, E denotes epithelial state, M mesenchymal state, I the intermediate state in the three-state EMT system, and 1, 2, and 3 denote the three intermediate states I_1, I_2, and I_3 in the four-state and five-state systems. The parameter k_{sI} is the self-renewal rate of the intermediate population in the three-state EMT system while the parameters k_{s1}, k_{s2}, and k_{s3} are the self-renewal rates of the I_1, I_2, and I_3 populations in the five-state process.

The sensitivity of each parameter is measured by the average change in the average coefficient of variation upon perturbation by 50%. For some transition rates, a positive change in the parameter value due to perturbation results in a significantly smaller average coefficient of variation but this parameter value can be discounted because it lies outside the designated value range that we used in our optimization algorithm Figures 3.4 and 3.8.

3.2 Results

3.2.1 Intermediate states reduce stochastic fluctuations in EMT

We first asked if the intermediate state helps to attenuate the fluctuations in E and M cells. We used two optimized parameter sets for a two-state (E and M) model and a three-state (E, I and M) model for comparison, and we obtained two representative trajectories under the influence of identical level of input noise in E and M states. When we compared these two trajectories in terms of the fractions of M cells (Figure 3.2A), we found that the fluctuations

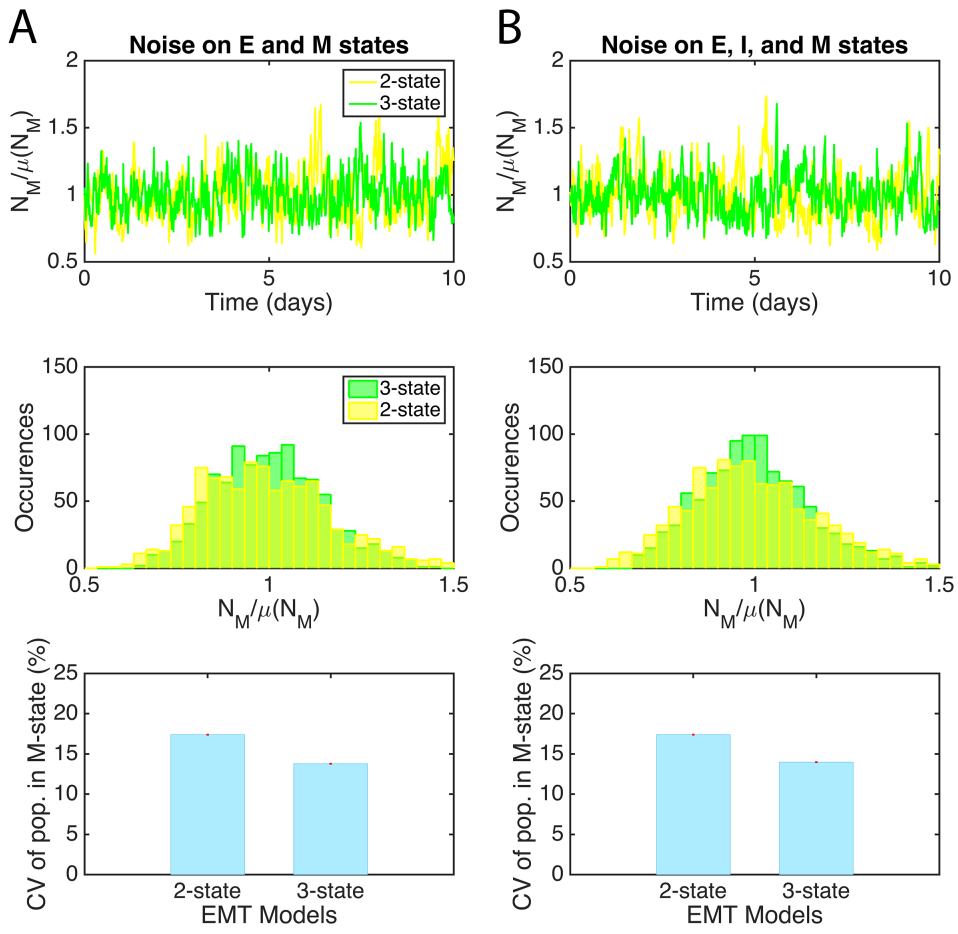

Figure 3.2: Comparison of the noise attenuation property between the two-state and three-state EMT systems in terms of the effects on the mesenchymal population size. **A)** Multiplicative noise is introduced to the epithelial and mesenchymal populations. **B)** Multiplicative noise is introduced to the epithelial, intermediate (three-state system), and mesenchymal populations. Top two panels are the time-course trajectories that represent the normalized number of mesenchymal cells $N_M/\mu(N_M)$ over a period of 10 days. To obtain the normalization, we perform stochastic simulations on the steady state population of mesenchymal cells, then divide the mesenchymal population size at each time point by its mean value obtained over the 10-day period. In the middle two panels, the normalized mesenchymal population size is plotted against the number of times that particular size occurs. Yellow: two-state EMT, green: three-state EMT. Bottom two panels display the quantification of the noise attenuation performance of the two-state and three-state systems using the mean and standard deviation of the coefficients of variation (CV) of the mesenchymal population. The mean is plotted here in the form of a bar graph (blue), while the standard deviation is described by the red error bar.

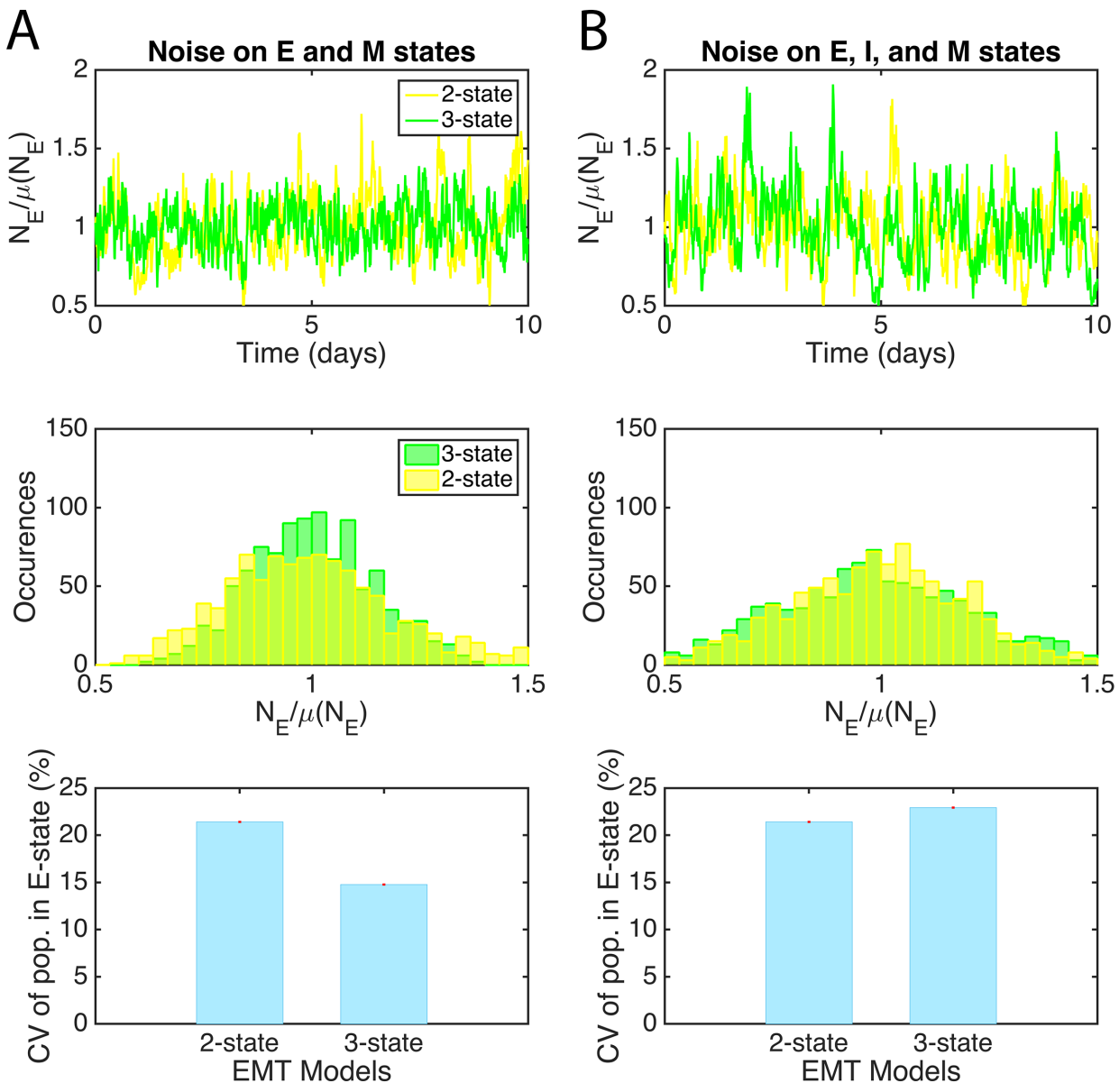

Figure 3.3: Comparison of the noise attenuation property between the two-state and three-state EMT systems in terms of the effects on the epithelial population size. **A)** Multiplicative noise is introduced to epithelial and mesenchymal populations. **B)** Multiplicative noise is introduced to the epithelial, intermediate (three-state system), and mesenchymal populations. Top two panels are the time-course trajectories that represent the normalized number of epithelial cells $N_E/\mu(N_E)$ over a period of 10 days. Middle two panels illustrate the distribution of the different population sizes of the normalized epithelial population. Here, the normalized epithelial population size is plotted against the number of times that particular size occurs. The color coding scheme for each system is similar to that of Figure 3.2. Bottom two panels display the quantification of the noise attenuation performance of the two-state and three-state systems.

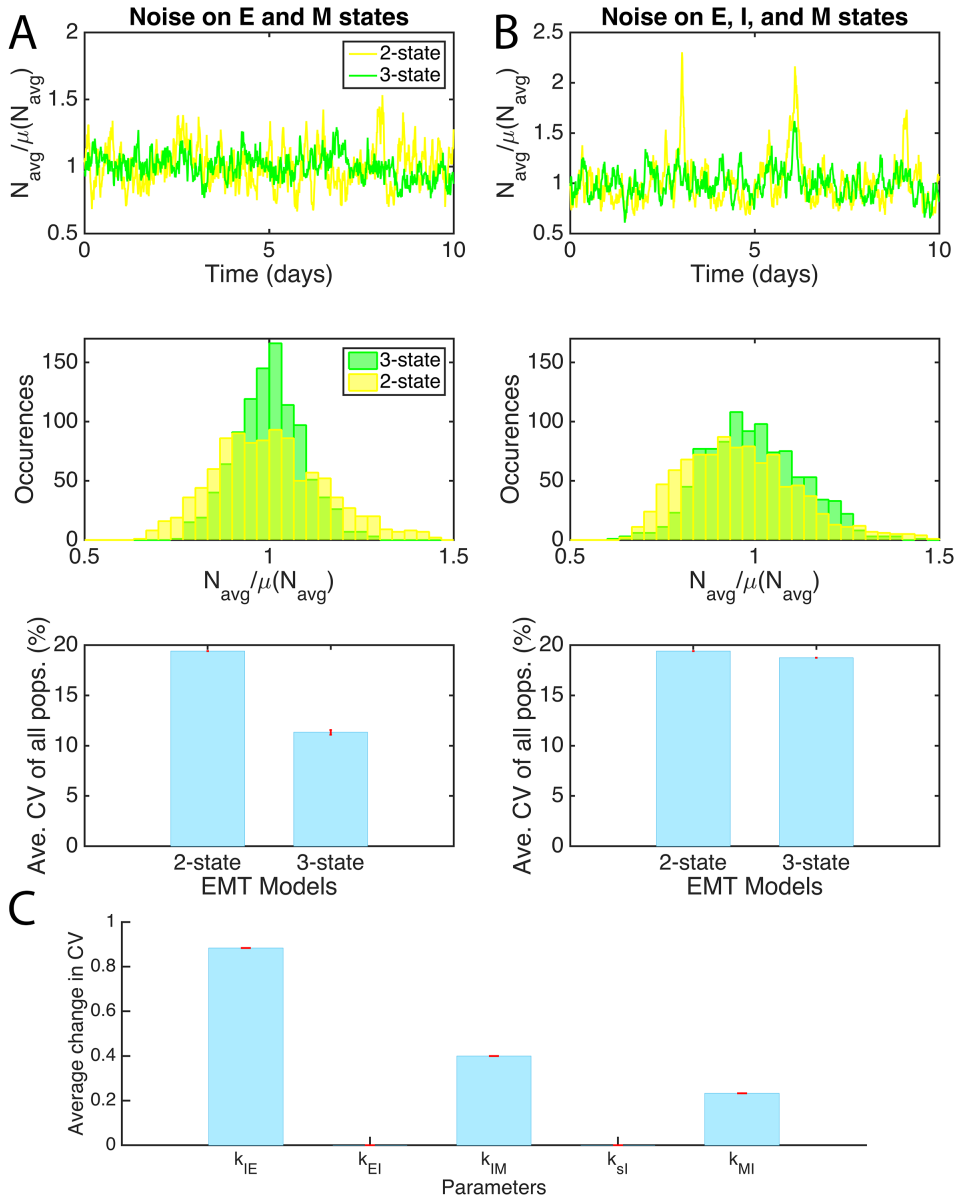

Figure 3.4: Comparison of the noise attenuation property between the two-state and three-state EMT systems on the average population size. **A)** Noise is introduced to epithelial and mesenchymal populations. **B)** Noise is introduced to the epithelial, intermediate (three-state system), and mesenchymal populations. Top panels: trajectories of normalized average number of cells $N_{avg}/\mu(N_{avg})$ over 10 days. Middle panels: distribution of the different sizes of the normalized average population using the color coding scheme of Figure 3.2. Bottom panels: quantification of the noise attenuation performance of both systems. **C)** Sensitivity analysis of the parameters representing unique cell transition rates in the three-state EMT system. Here, we plot the mean of the average change in the average CV as bar graphs (blue) accompanied by red error bars that describe the standard deviation of the average change.

in the fraction of M cells were reduced by 3.6% after adding the intermediate state. To exclude the possibility that the reduction of fluctuations is due to the lack of input noise in the intermediate state, we compared the two-state model and the three-state model with noise in all of the three states, and the latter model still showed discernible reduction of fluctuations in fraction of M cells under this condition (Figure 3.2B). We performed similar analysis on the fraction of E cells, and we found that when the input noise was only in E and M states, the three-state model performed significantly better than the two-state in terms of minimizing the fluctuations in E cells (Figure 3.3A), but when the input noise was in all the states, adding intermediate state resulted in a moderate increase of the fluctuations (Figure 3.3B). These results suggest that the intermediate state attenuates the fluctuations in M cells, and it may decrease or increase the fluctuations in E state depending on the source of noise.

We quantified the average fluctuations of all cellular states with the two-state and the three-state models. When we added noise on E and M states, the average cellular fluctuations were markedly reduced in the presence of the intermediate state (Figure 3.4A), and there was still a moderate reduction of average fluctuations when the noise was on all cellular states (Figure 3.4B). This suggests that the overall performance of the three-state model is better than that of the two-state model in terms of stability of phenotypic compositions.

We next asked which cellular state transition rates are critical for maintaining the stability. By perturbing the parameters representing the rates transitions among the three cellular states (input noise was added to all three states), we found that the performance of noise attenuation in the three-state system is most sensitive to three parameters: the transition rate from I to E (k_{IE}), the transition rate from I to M (k_{IM}) and the transition rate from M to I (k_{MI}) (Figure 3.4C). These results suggest that the transitions involving the intermediate states are important for noise attenuation, and they further corroborate the advantage of the intermediate state. They correlation between these transition rates and the ability of the system to attenuate noise may be tested in future experiments.

3.2.2 Multiple intermediate states enhance noise attenuation

Next, we explored the advantages of modeling the EMT system with multiple steady states. Using identical input noise on E and M populations, we examined the noise buffering property of three different EMT models: three-state, four-state, and five-state transitions. Analysis of the fluctuations in M cells demonstrates a reduction in noise effects in systems with more than one intermediate states, as evident by the normalized trajectories of the fraction of M cells and the coefficient of variation that reflects the variability of the mesenchymal population size (Figure 3.5A). In addition, five-state system filters noise more efficiently than four-state system. Similarly, when we introduced noise to one intermediate population in addition to the E and M populations, we observed the same correlation between the number of intermediate populations in an EMT system and the noise attenuation property of that system (Figure 3.5B). We also confirmed this result by adding noise to all the populations in every EMT system, where the five-state system achieved the best noise filtering results, followed closely by the four-state system (Figure 3.5C). From analyzing the responses of the average intermediate population to noise with similar approach, we found that the populations of the five-state and four-state systems outperformed that of the three-state system in attenuating noise at the intermediate levels, regardless of how noise was introduced to the different populations (Figure 3.6A-C). This suggests that the number of stem cell sub-states correlates with their ability to reduce the fluctuations of stem cell population. When we studied the effects of noise on the fraction of E cells, we did not observe a similar trend to that of the M and I cells. When input noise was added to E and M populations only, the epithelial population of the five-state system was more susceptible to the fluctuations (Figure 3.7A) while the epithelial population of the same system showed the least variability when noise was also added to one intermediate population (Figure 3.7B). With noise in every state,

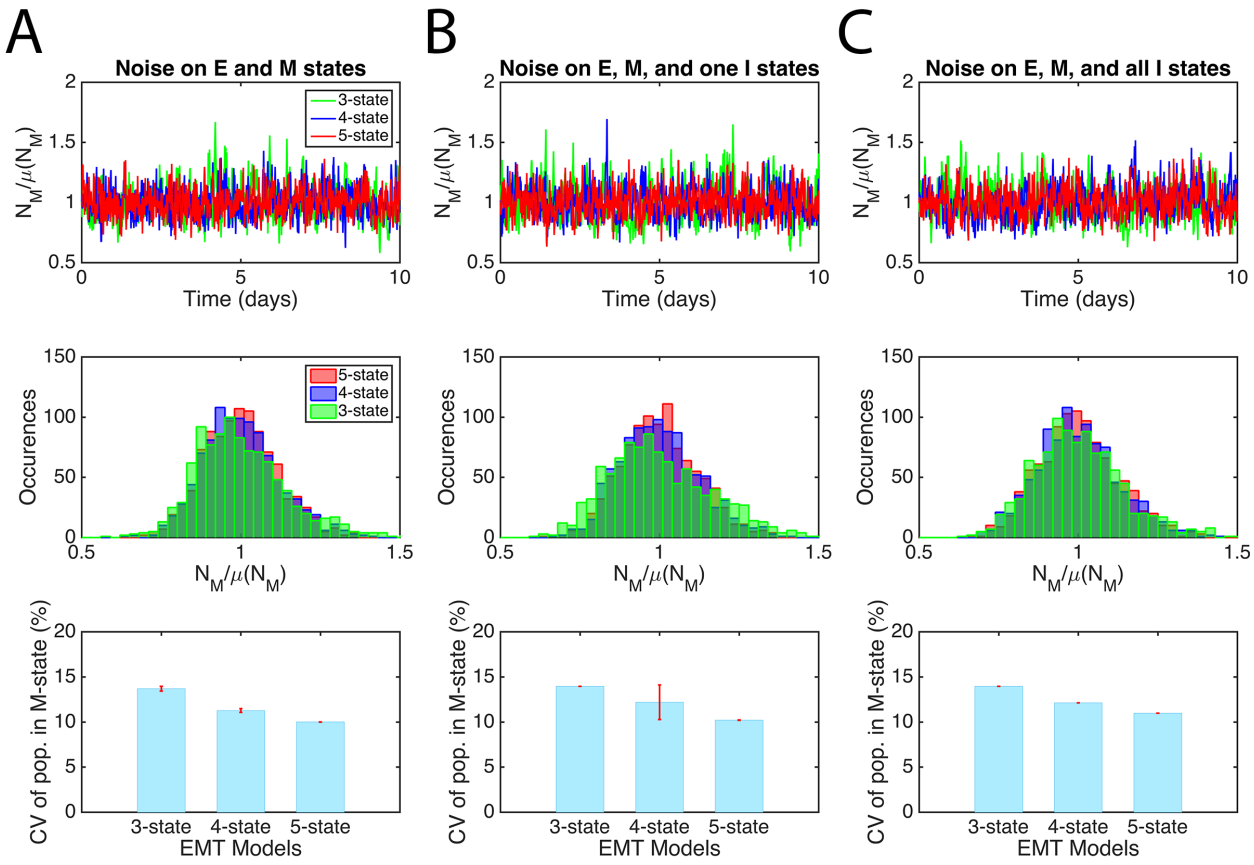

Figure 3.5: Comparison of the noise attenuation property between the three-state, four-state, and five-state EMT systems in terms of the effects on the mesenchymal population size. **A)** Multiplicative noise is introduced to epithelial and mesenchymal populations. **B)** Multiplicative noise is introduced to the epithelial, mesenchymal, and one intermediate populations. **C)** Multiplicative noise is introduced to all the populations. Top three panels are the time-course trajectories that represent the normalized number of mesenchymal cells $N_M/\mu(N_M)$ over a period of 10 days. To obtain the normalization, we perform similar stochastic simulations to those in Figures 3.2-3.4. Middle three panels illustrate the distribution of the different population sizes of the normalized mesenchymal population. Here, the normalized mesenchymal population size is plotted against the number of times that particular size occurs. Green: three-state EMT, blue: four-state EMT, red: five-state EMT. Bottom three panels display the quantification of the noise attenuation performance of the three-state, four-state, and five-state systems.

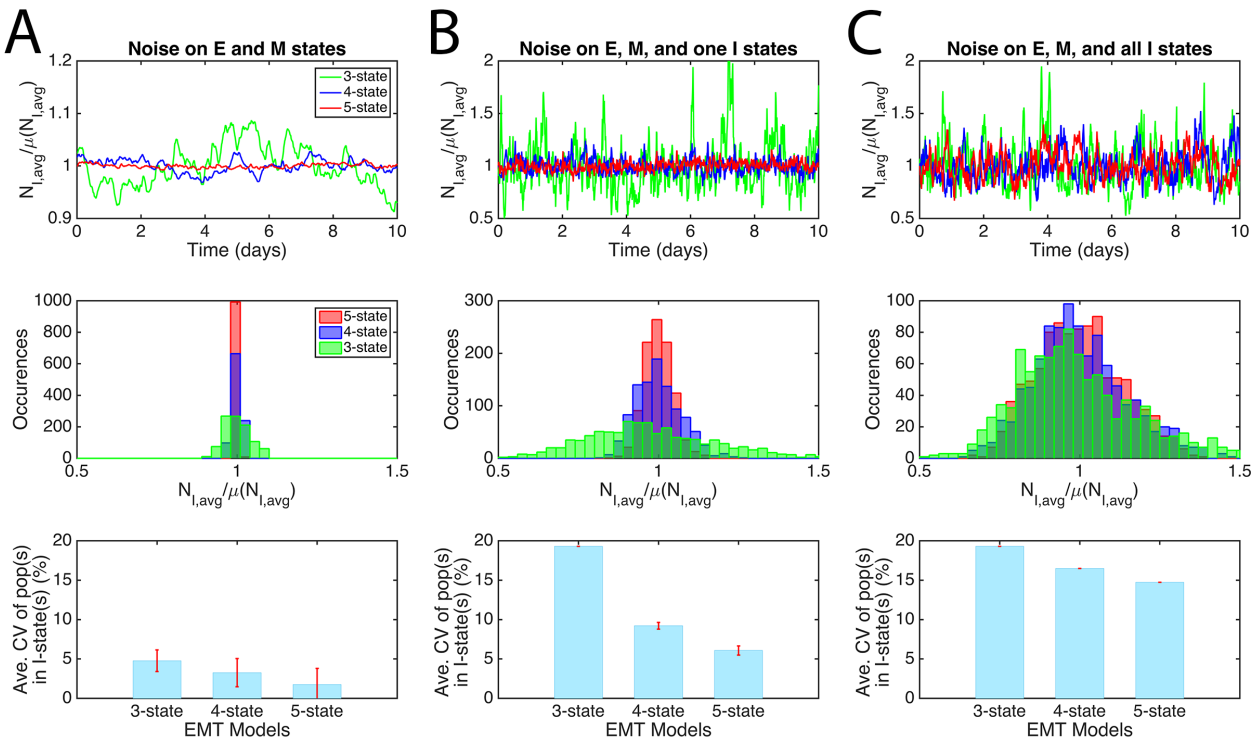

Figure 3.6: Comparison of the noise attenuation property between the three-state, four-state, and five-state EMT systems in terms of the effects on the average intermediate population size. **A)** Multiplicative noise is introduced to epithelial and mesenchymal populations. **B)** Multiplicative noise is introduced to the epithelial, mesenchymal, and one intermediate populations. **C)** Multiplicative noise is introduced to all the populations. Top three panels are the time-course trajectories that represent the normalized average number of cells taken over all the intermediate states $N_{I,avg}/\mu(N_{I,avg})$ over a period of 10 days. Middle three panels illustrate the distribution of the different population sizes of the normalized average intermediate population. Here, the normalized average intermediate population size is plotted against the number of times that particular size occurs. The color coding scheme for each system is similar to that of Figure 3.5. Bottom three panels display the quantification of the noise attenuation performance of the three-state, four-state, and five-state systems.

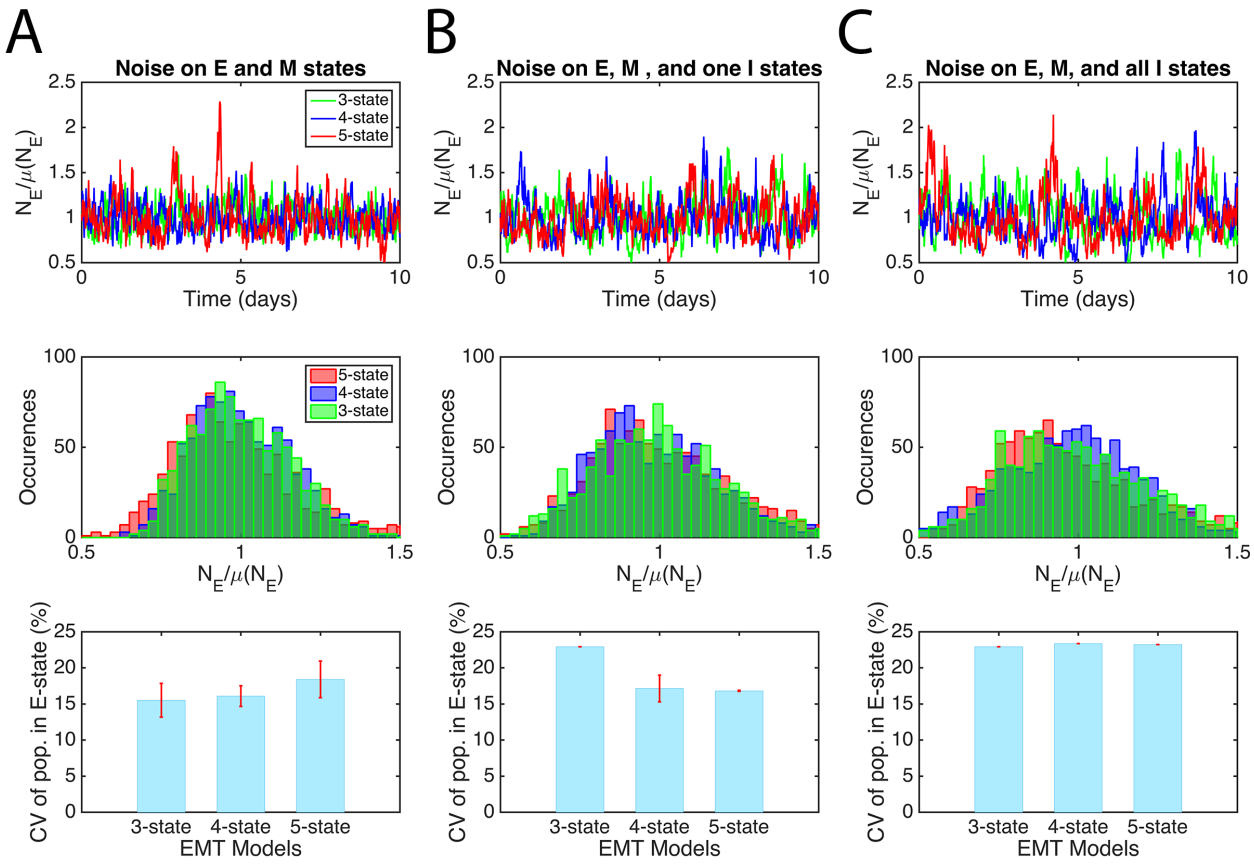

Figure 3.7: Comparison of the noise attenuation property between the three-state, four-state, and five-state EMT systems in terms of the effects on the epithelial population size. **A)** Multiplicative noise is introduced to epithelial and mesenchymal populations. **B)** Multiplicative noise is introduced to the epithelial, mesenchymal, and one intermediate populations. **C)** Multiplicative noise is introduced to all the populations. Top three panels are the time-course trajectories that represent the normalized number of epithelial cells $N_E/\mu(N_E)$ over a period of 10 days. Middle three panels illustrate the distribution of the different population sizes of the normalized epithelial population. Here, the normalized epithelial population size is plotted against the number of times that particular size occurs. The color coding scheme for each system is similar to that of Figure 3.5. Bottom three panels display the quantification of the noise attenuation performance of the three-state, four-state, and five-state systems.

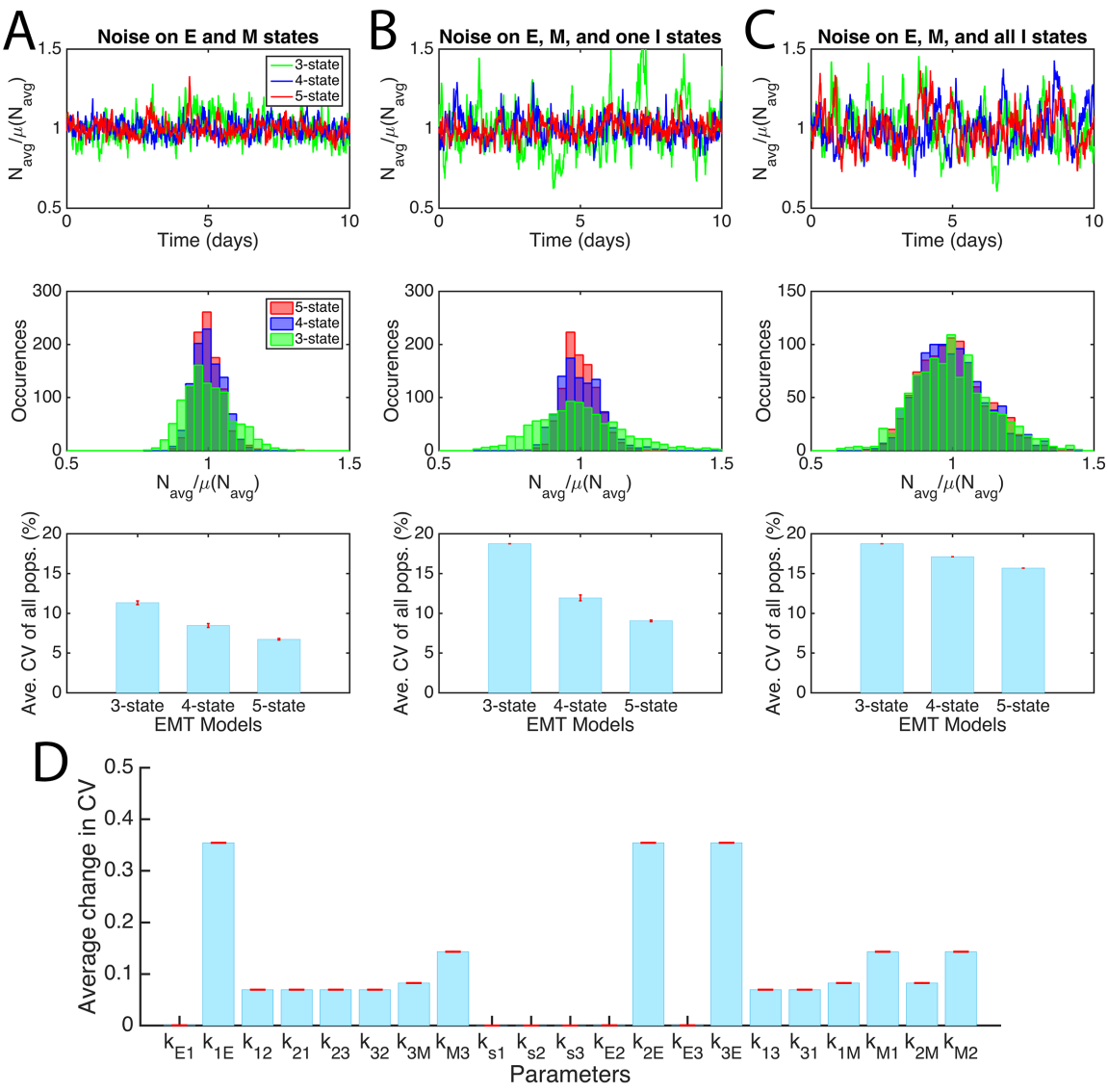

Figure 3.8: Comparison of the noise attenuation property between the three-state, four-state, and five-state EMT systems in terms of the effects on the average population size. **A)** Multiplicative noise is introduced to epithelial and mesenchymal populations. **B)** Multiplicative noise is introduced to the epithelial, mesenchymal, and one intermediate populations. **C)** Multiplicative noise is introduced to all the populations. Top three panels are the time-course trajectories that represent the normalized average number of cells $N_{avg}/\mu(N_{avg})$ over a period of 10 days. Middle three panels illustrate the distribution of the different population sizes of the normalized average population. The color coding scheme for each system is similar to that of Figure 3.5. Bottom three panels display the quantification of the noise attenuation performance of all the EMT systems. **D)** Sensitivity analysis of the parameters representing unique cell transition rates in the five-state EMT system. Here, we plot the mean of the average change in the average CV as bar graphs (blue) accompanied by red error bars that describe the standard deviation of the average change.

the difference in the performance of each system in filtering noise at the epithelial level was not discernible (Figure 3.7C). These results suggest that when the number of intermediate state varies, there might be a tradeoff between attenuating noise in E cells and other cells. Lastly, we quantified the overall performance of each EMT system by considering the fluctuations on the average population of all cellular states. In the scenario where noise is only introduced to the E and M states, the five-state system performs the best at suppressing noise, while the four-state system displays better noise attenuation than the three-state system (Figure 3.8A). With the addition of noise in one intermediate state, the four-state and the five-state EMT processes demonstrate improvement in noise attenuation from their three-state counterpart by 6.9% and 9.6% respectively (Figure 3.8B). Likewise, we confirmed our results with the addition of noise to all cellular populations and concluded that the EMT system achieves better noise attenuation property when more intermediate states are taken into consideration (Figure 3.8C).

Using the five-state EMT process with noise in every state, we analyzed the sensitivity of the parameters that described the cellular transition/self-renewal rates. For our analysis, we examined the average change in the average coefficient of variation of all five populations upon perturbing each parameter. We performed the optimization procedure ten times, whereupon, we carried out the perturbations each time and presented our results in Figure 3.8D. We found that the parameters that described the transition rate from any intermediate population I_1, I_2, or I_3 to the epithelial population (k_{1E}, k_{2E}, k_{3E}) are most sensitive to their own perturbations, resulting in a marked increase in the average coefficient of variation, therefore a decrease in noise buffering ability of this EMT system. Besides those three transition rates, the cellular transitions from the mesenchymal population to the intermediate populations (k_{M1}, k_{M2}, k_{M3}) , also show considerable sensitivity to perturbations. We thus recognize the crucial roles that the intermediate states such as the I_1, I_2, and I_3 states play in ensuring the better noise attenuation performance of the five-state system. The significance of the transitions between intermediate states and the others reinforce the notion that having

multiple intermediate states is beneficial to reducing the effects of the fluctuations on the overall population.

3.3 Discussion

Studying the multi-step and reversible dynamics of EMT is essential for understanding the roles of EMT in various biological processes such as development and cancer progression. Previous studies revealed multiple intermediate states lying between terminal epithelial and mesenchymal states [40, 42, 85, 99], and this is consistent with the observations that epithelial cell populations show remarkable heterogeneity in normal and tumor tissues [59, 113]. These intermediate states are associated with stemness and invasiveness during cancer progression [29, 49], but the functions of these multiple intermediate states and the complex transitions among all the phenotypic states are unclear. In this study, we used mathematical models to show that systems with multiple intermediate (stem cell) states have advantages in attenuating fluctuations of the heterogeneous cell population, and this effect is particularly beneficial for maintaining stable fractions of stem cells in cell populations, thus providing a strategy for maintaining homoeostasis in facing stochasticity of cell fate changes. The possible performance objective at systems level suggests a new design principle for multi-step EMT and other multi-state systems involving transitions among diverse types of cells.

Complex networks in biological systems have been extensively studied in terms of design principles and performance objectives [66, 101]. However, most of these studies focused on intracellular networks of molecular interactions and influence, and much less is known for networks formed by phenotypic transitions among cell types [34]. We showed that the later type of networks can carry rich information in terms of performance objectives such as facilitating noise attenuation. This finding improves our understanding of systems design at

tissue level, and suggests that the formation of cellular states and the transitions among them may have evolved to reach performance objectives that are beneficial to the organisms. It is likely that multiple performance objectives are influencing the evolution of the cellular networks, and further studies are needed to identify these objectives.

Due to the stochastic nature of biochemical reactions and fluctuating environmental conditions, many biological systems are designed to attenuate noise such that stable response to signals can be attained [15, 105]. Intermediate states and regulators have been previously found to improve the robustness of developmental patterning [10]. Our finding that adding intermediate cell states can reduce the fluctuations in cell population provides another strategy for noise attenuation at tissue level. The fluctuations that we included in our simulations represent the stochasticity of cell division [98], cell death [26] and phenotypic transitions [13], all of which were previously observed in experiments. Nonetheless, the sources of these types of noise are mostly molecular fluctuations that affect various types of rates of the cellular activities [118]. In order to describe the dynamics of the noise in more details, future models will need to incorporate the stochastic dynamics of biomolecules into the framework.

Multiple intermediate cell states may arise from complex gene regulatory networks with interconnected positive feedback loops [40, 39]. It is conceivable that the formation of more intermediate cells would require more complex gene regulatory networks, which in turn need some other strategies and/or more energy to control. Therefore, the correlation between the number of states and the ability to attenuate noise suggests that a tradeoff between the simplicity of the gene regulatory network and noise attenuation may exist when the system is subject to design via evolution.

Dynamic equilibrium of sub-states of stem cells is observed not only in the EMT system, but also in hematopoietic and embryonic stem cell populations [36, 92, 109]. Moreover, it was suggested that heterogeneity of cell population might be related to disseminated cancer cell dormancy [94]. Although these sub-states of stem cells might represent distinct functional entities in specific contexts, the recurring phenomena indicate that there might be

advantages of this dynamic behavior at systems level. Our modeling work provides a general framework to study the cell differentiation systems with multiple sub-states of stem cells, and our conclusions with respect to the noise attenuation property of multiple stem cell states can apply to other systems, such as embryonic stem cell and disseminated cancer cell [36, 94]. In addition, feedback control among multiple cellular states have been shown to be critical to the growth and stability of cell populations [62, 68, 72]. Future works are warranted to investigate the role of feedback controls involving the multiple intermediate states.

In conclusion, our computational analysis shows that the existence of multiple intermediate states between epithelial and mesenchymal states and the transitions among these states are able to attenuate the fluctuations of the fractions of cell population. We found a general correlation between the number of states and the ability to attenuate fluctuations. In particular, the fluctuations of the stem cell populations are reduced by increasing the number of intermediate states. These results improve the understanding of the intermediate states in the EMT system in terms of performance objectives, and provide insights into the stem cell systems with multiple sub-states in general.

Chapter 4

Multiscale Model of a Heterogeneous Population in the EMT Spectrum

4.1 Model Construction

4.1.1 Construction of multiscale model

We constructed a multiscale model to monitor the gene expression levels over time of key proteins that are featured in our EMT gene regulatory network. Then we observed the growth of the cell population of each phenotype under stochastic influences over time. In our model, we made the following assumptions:

1. The initial population is composed of 100 cells, where 70% of the cells are epithelial, 10% I_1, 10% I_2, and 10% mesenchymal. Our inferences on the population proportions of all the cell types are based on observations of developing mammary epithelial tissues during puberty and pregnancy, with the majority of the cells being epithelial.

Parameter	*Description*	*Value*
r_i	Intrinsic cell cycle length of a non-stem cell	$r_i \sim \mathcal{N}(700, 200)$
τ_i	Intrinsic cell cycle length of a stem cell	$r_i \sim \mathcal{N}(500 * k, 50)$
k	Sum of the normalized concentrations of *Zeb*1 mRNA and Ovol2 protein of the cell	Varies
d	Death rate of TD cells	$d \sim \mathcal{N}(1000, 100)$
M	Number of times a non-stem cell divides before becoming TD cell	$M \sim \mathcal{U}(2, 7)$

Table 4.1: List of parameter values used in the lineage dynamics

2. All cells divide at a normally-distributed rate. Every time a cell divides, it passes the gene expression levels of all the EMT factors as initial conditions to its daughter cells. We compare the gene expression levels of each daughter cell at equilibrium to the expression levels of different steady states in the EMT spectrum to determine the eventual cell fate of the daughter cell.

3. Stochastic effects are integrated into our model by adding two types of noise, whose details will be elaborated on in the next section.

4. We ascribe stemness to cells in the I_1 state. Stem cells can either undergo asymmetric division to give rise to a stem cell and a non-stem cell or symmetric division to give rise to two stem cells or two non-stem cells. Stem cells do not die and can divide infinitely.

5. All other types of cells including epithelial and mesenchymal cells eventually become terminally differentiated (TD) cells. They divide a finite number of times N before becoming TD cells. TD cells die at a normally-distributed rate d in Table 4.1.

We performed simulations of our multiscale model over a time span of seven average cell division cycles. A comprehensive diagram of our multiscale model is included in Figure 4.1.

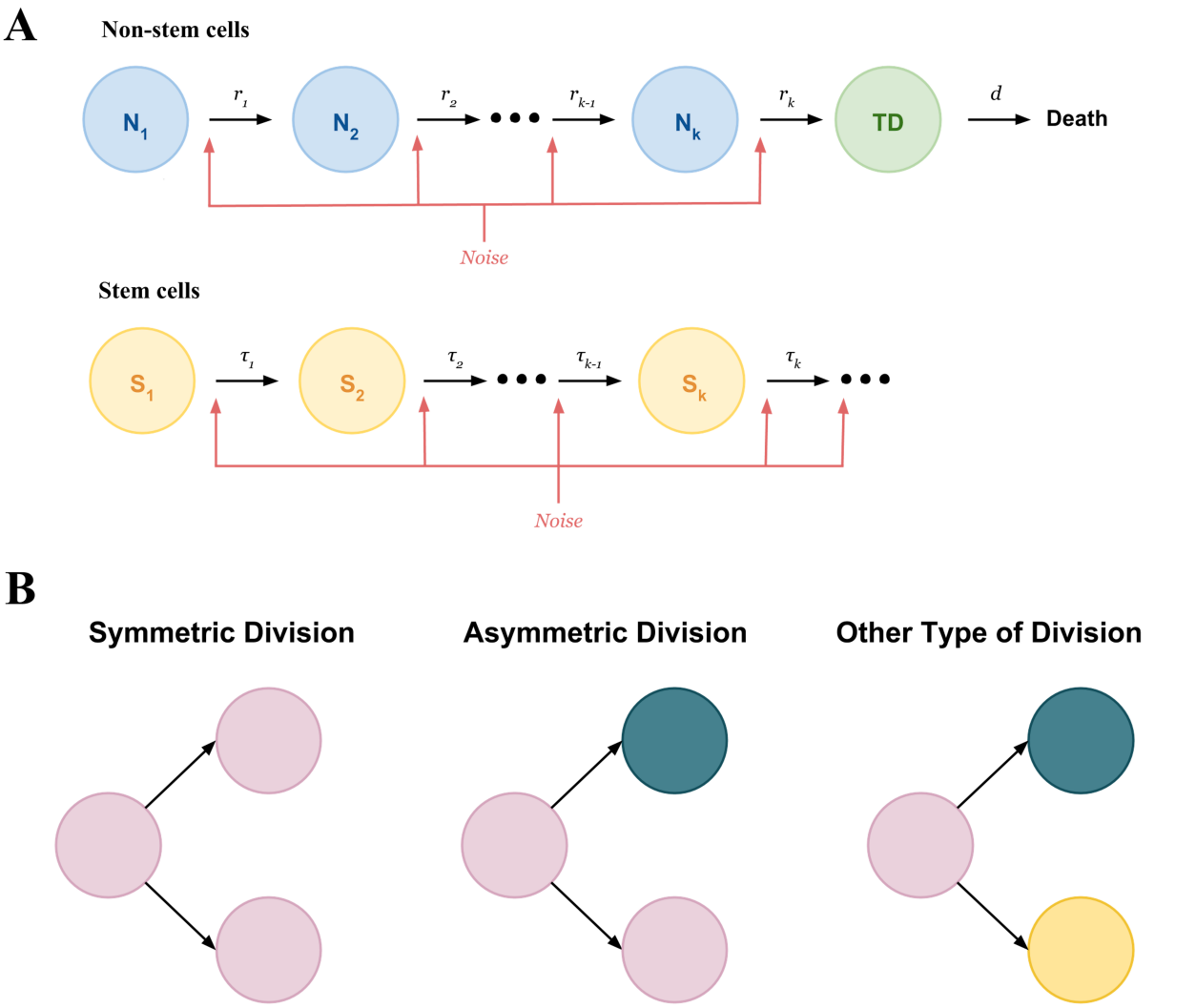

Figure 4.1: **A)** Depiction of the multistage lineage dynamics involving non-stem cells (E, I_2, and M) and stem cells (I_1). Non-stem cells divide a finite number of times before becoming TD cells and die at a rate d while stem cells can divide indefinitely. r_i is the intrinsic cell cycle length for non-stem cells and τ_i is the intrinsic cell cycle length for stem cells. The values of the death rate and cell cycle lengths are normally distributed. For parameter values, review Table 4.1. Noise is introduced every time a cell divide. **B)** Schematic depiction of the possible types of division. From left to right: a cell of type T can divide symmetrically into two T cells to proliferate or it can also divide asymmetrically to produce one T cell and one cell of a different type. Lastly, noise can induce T cell to produce two daughter cells that are not T. These daughter cells can be of the same or different types.

4.1.2 Experimentation with two different types of noise

In order to incorporate cell fate reprogramming to our multiscale model, we introduce various types of noise to our EMT system independently. We first constructed noise by perturbing the gene expressions of the mother cell upon its division into two daughter cells. In this scenario, the perturbed expressions serve as the initial conditions for the simulations of the daughter cells. We assume that the aggregate concentration of each gene in the daughter cells remains unchanged from the respective gene's concentration in the mother cell. Hence, for each gene, the sum of the initial expressions of the daughter cells should be equal to two times the amount of gene expression in the mother cell. From here on, we will refer to this type of division noise as Type 1. We next introduce a different type of noise, which we will refer to as Type 2. For Type 2 noise, we applied multiplicative noise to a few parameters that model the key proteins of the EMT circuit upon cell division, including the concentration of exogenous TGFβ level, the transcription rate of *Snail1* mRNA, and the production rate of miR-200, as described in Table 4.2.

We further experimented with Type 1 noise in two different ways. First, we simulated gene fluctuations by allowing all gene expressions to simultaneously increase in one daughter cell while decreasing these expressions by each gene's corresponding amount in the second daughter cell. In the second situation, we allowed the changes in gene expressions to be uncorrelated in each daughter cell, i.e one gene can receive an enhancement in its expression while another gene undergoes a reduction. We compare the effects on cell population from these different fluctuations in Figures 4.2 and 4.3.

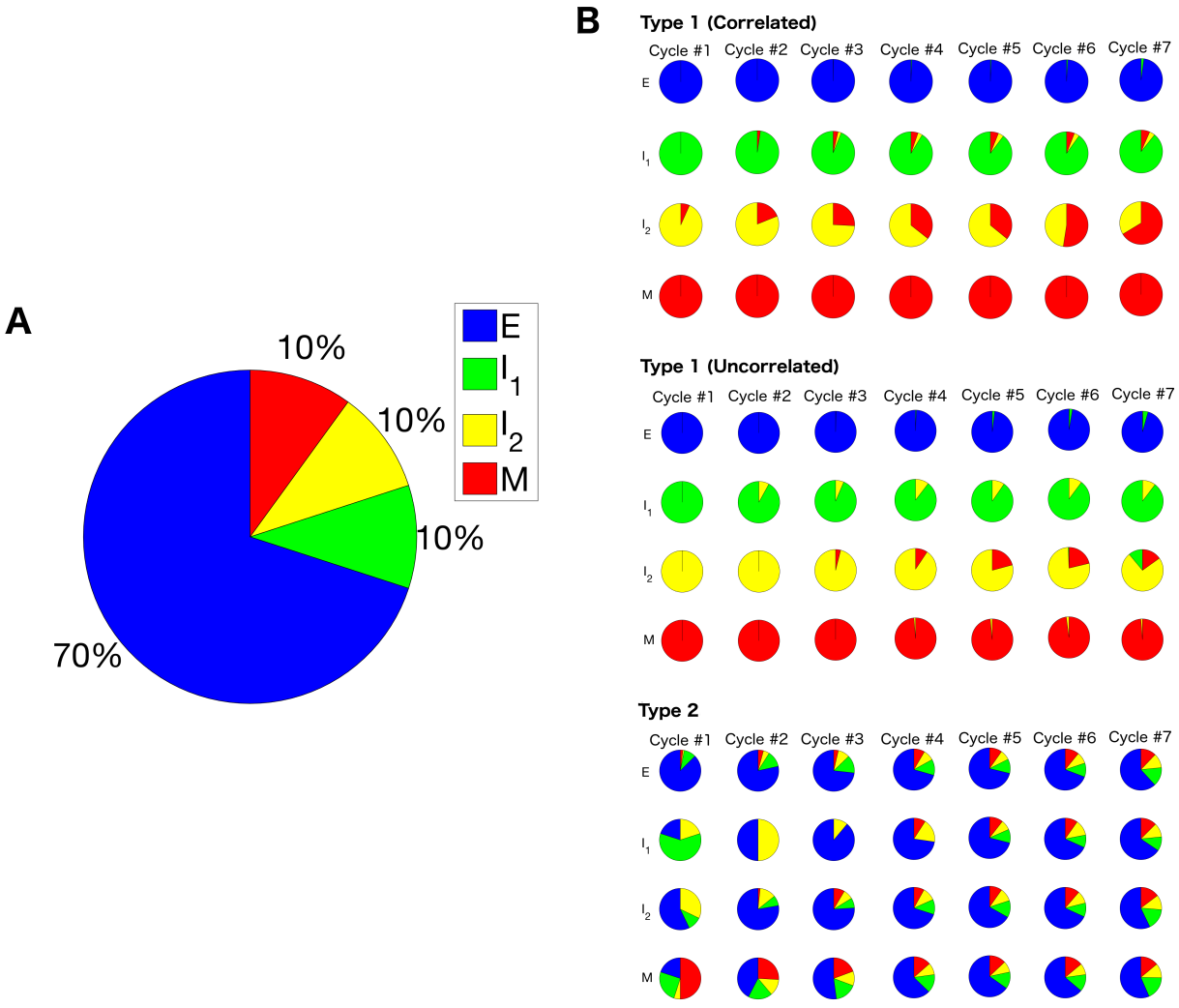

Figure 4.2: **A)** The distribution of the population at the initial stage (initial conditions) in terms of its phenotypic composition. **B)** The composition of each phenotype's population in terms of the phenotypes of its ancestors at the end of every cell cycle. The color scheme to indicate the different phenotypes of the ancestor cells is depicted in (A). The proportion of cells that is originated from E state (epithelial cells) is colored blue. The proportions of cells that are originated from I_1 and I_2 states (two intermediate states) are colored green and yellow respectively. The proportion of cells that is originated from M state (mesenchymal cells) is colored red. Each figure in (B) corresponds to a different type of noise. The results from this figure are obtained from one simulation.

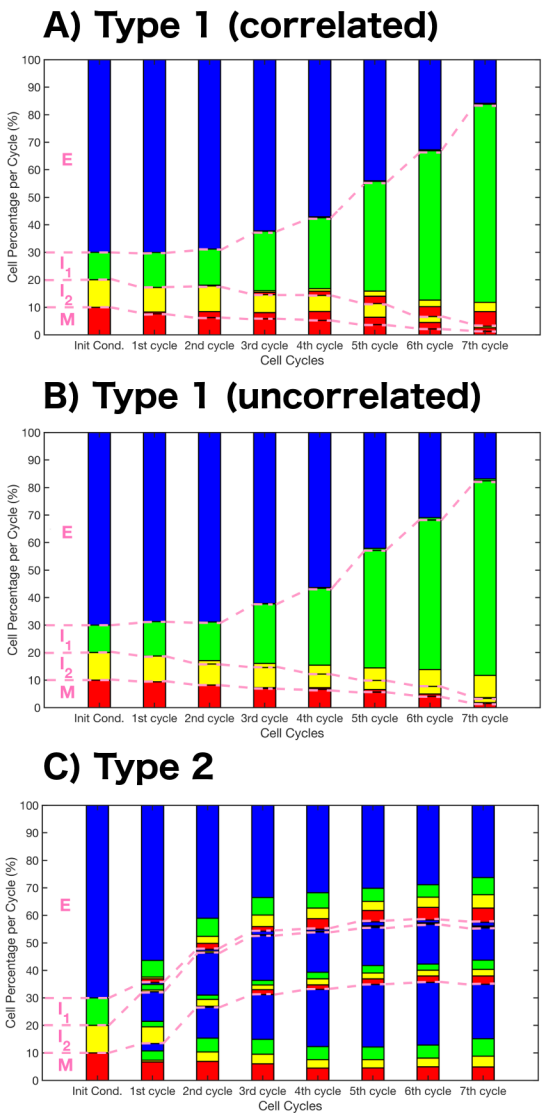

Figure 4.3: The distribution of the total cell population at the end of every cell cycle as well as the distribution of each phenotypic fraction of that population in terms of the phenotypic composition of each fraction's ancestors. The pink dashed lines separate the different proportions of the total population according to their appropriate phenotypes denoted by pink letters on the left margin of each figure. Within each proportion (between two pink dashed lines), the composition of that portion of the population is further broken up into smaller fractions according to the phenotypes of the original ancestors (from the initial conditions). These fractions share the same color-coding scheme as in Figure 4.2: E blue, I_1 green, I_2 yellow, and M red. The bar at the very left represents the initial makeup of the original population, where 70% of the cells are epithelial, 10% I_1, 10% I_2, and 10% M. All the results are obtained from one simulation of cellular activities for seven cell cycles. **A)** The distribution of the cell population obtained from Type 1 correlated noise. **B)** The distribution of the cell population obtained from Type 1 uncorrelated noise. **C)** The distribution of the cell population obtained from Type 2 noise.

Noise Type	*Noise term*	*Value*
Type 1 (correlated)	$n_g * g$ n_g: noise coefficient, g: gene	$n_g \sim \mathcal{N}(0, 0.7)$ and $\|n_g\| <= 0.9$ $sgn(n_g)$ is the same for all genes g
Type 1 (uncorrelated)	$n_g * g$	$n_g \sim \mathcal{N}(0, 0.7)$ and $\|n_g\| <= 0.9$
Type 2	$n_p * p$ n_p: noise coefficient, $p = TGF0, k_{200}, k_{snail}$ Values of p can be found in [40]	$n_p \sim \mathcal{N}(0, 0.7)$ and $\|n_p\| <= 0.9$

Table 4.2: List of the parameters used to construct two types of noise

4.2 Results

4.2.1 Type 1 and Type 2 noises exert distinctive effects on cell fate reprogramming

In Figure 4.2, the singular pie on the very left describes the initial heterogeneous distribution of the cell population where 70% of the population are cells of epithelial phenotype and 10% of I_1, I_2, and mesenchymal phenotypes respectively. Each of the phenotype is color-coded in the initial condition pie and we will adhere to this color-coding scheme for the rest of the figure. The subsequent columns of pie charts are ordered by the division cycle time, where the first column indicates the cell population at the end of the first division cycle and so on. Each row of the figure corresponds to the aforementioned phenotypes indicated by the letter E, I_1, I_2, and M. Each of the smaller pies represents the distribution of the cells of an indicated phenotype at a specific cell division cycle according to the phenotypes of the original ancestral cells from the initial conditions. For example, at the end of the second division cycle, the I_1 cells are either descended from I_1 cells or M cells from the initial population, with the majority of these cells descending from the initial I_1 population, as indicated by a major portion of the pie being colored green, Figure 4.2B Type 1 (correlated). We note that Figure 4.2 corresponds to one simulation of cellular activities that spans a period of seven

average cell division cycles.

When the changes in gene expressions are correlated in each daughter cell, we only observe cell conversion in the MET direction, i.e there exists no instance where a cell of the initial population gives rise to a descendant cell that belongs to a more mesenchymal phenotype than that of its initial ancestor, which can be generalized to the lack of conversion in the EMT direction between any two types of cells. However, if these gene fluctuations are uncorrelated, we observe some instances of cell conversion in EMT direction. For example, starting at the end of the sixth division cycle, we observe that there is a portion of I_2 cells whose ancestors belong to the I_1 phenotype, indicating that there must be some conversion towards a more mesenchymal phenotype at some point during the course of the simulation. We suspect that correlated noise only favors the MET direction. Although both EMT and MET agents in the regulatory network undergo an increase (decrease) in gene expressions, the MET agents have more influence on the network and consequently dictate the direction of the cell fate of the daughter cells. Allowing the fluctuations to be uncorrelated mitigates the effects of the MET agents' influence. Between the correlated and uncorrelated Type 1 noises, correlated fluctuations yield relatively fewer cell state transitions than uncorrelated ones.

Meanwhile, we observe a more active differentiation in both EMT and MET directions for each cell phenotype throughout each cell cycle in the case of Type 2 noise, indicating that this type of noise is able to achieve a more heterogeneous cell population in terms of their origins. It is worth noting that the cellular population at the end of the seventh division cycle seems to stabilize at similar distributions between all the different phenotypes. This similarity in the contribution of all the different phenotypic populations to the final population of each cell type has important implications in understanding how the EMT and MET processes influence the growth of a cell population. If a delicate balance between the contribution of different types of population is indeed inherent to the EMT and MET processes irrespective of the final phenotype, then it will provide an explanation on how EMT and MET govern

cell proliferation. We have yet to compare these simulation results with experimental data but it is important to note that the simulation results displayed in this figure are consistent between different independent simulations for all types of noise.

Another interesting observation from Figure 4.2 is that the final distribution for each phenotype bears great resemblance to the initial population distribution, suggesting that the distribution of the initial population determines the distribution of the final population of each phenotype. It is unclear whether changing the initial population distribution will also result in the same equilibrium of final phenotypic contribution as observed here. It is thus crucial to explore various sets of initial conditions in the future to ascertain a definitive conclusion on the role of the initial population distribution in regulating the different proportions of cells in the final population for a particular cell type.

4.2.2 Type 1 and Type 2 noises exert different influences on the proliferative potential of cells and the resulting population

Figure 4.3 represents a comprehensive description of the distribution of cell population at the end of any cell cycle as well as the origins of each phenotypic proportion of that population. The proportions of cells marked by different phenotypes E, I_1, I_2, and M are separated by the pink dashed lines. The letters to the left of the initial conditions bar in the figure denote the states in the EMT spectrum that each region enclosed by the pink dashed lines belongs to. Similarly to Figure 4.2, the initial conditions bar describes the distribution of the initial population where 70% of the cells belong to the E state and 10% belong to I_1, I_2, and M states respectively. At any division cycle, we can trace the lineage of cells of a particular phenotype using the same color-coding scheme as before. For example, at the end of the fifth cycle in Figure 4.3B, the region of the bar that denotes I_1 cells is split into two parts: a big green region denoting I_1 state and a smaller yellow region denoting I_2 state. In other words, all the cells in the I_1 state at the end of this cycle are originated from either I_1

or I_2 cells from the initial conditions. This figure also provides information on how many percent of the population each phenotype occupies. At the end of the fifth cycle, the I_1 cells comprise approximately 45% of the total population, Figure 4.3B. Figure 4.3 is obtained using only one simulation of cellular activities, but the results remain consistent between different independent simulations.

We note that in the case of Type 1 noise, regardless of how fluctuations are included to the initial conditions of the daughter cells, the I_1 phenotype quickly dominates the total population of the cells, comprising almost 80% of all the cells by the end of the seventh cycle. We suspect that this type of noise cannot overcome the stem-cell property ascribed to the I_1 phenotype. Even though many I_1 mother cells give rise to epithelial cells, the majority of them still retain their I_1 status and hence, stemness, and proliferate quickly to dominate the total population. In the mean time, the proportion of cells that belong to each of the other phenotypes declines over time. For the remaining analysis, we define ancestral cells to be the cells from the initial conditions. We observe that for each phenotype, most of the cell population is originated from ancestral cells of the same phenotype per each cycle while the majority of the remaining cells come from the immediately adjacent phenotype in the EMT direction. For example, at every division cycle, most of the cells of the epithelial phenotype come from ancestral cells that are epithelial, while the remaining cells trace their origins back to the adjacent phenotype in the EMT direction, which is I_1 state in this scenario. The main characteristic that distinguishes Type 1 correlated noise from uncorrelated noise is that a significant proportion of I_1 cells can trace their origin back to mesenchymal cells starting with fourth cycle.

Unlike Type 1 noise, Type 2 noise results in a discernible decrease of I_1 cells, i.e stem cells, signifying that Type 2 noise does not favor the stem-cell phenotype and causes significant fluctuations such that most of I_1 cells differentiate to other cell types. This subsequently affects the number of cells in the total population over time, as indicated by Figure 4.4. With a significantly smaller proportion of stem cells, the total population obtained from Type 2

noise is noticeably smaller than those obtained from the two cases of Type 1 noise. When we examine the other phenotypes besides I_1, we notice that the overall proportion of cells that carry E phenotype decreases after three division cycles and eventually stabilizes at around 40-45%. The proportions of both I_2 and M phenotypes increase then stabilize respectively during the seven cycles, with the proportion of mesenchymal cells increasing by a significant percentage. However, it is apparent that at every cycle, most of the total population traces its lineage back to an epithelial ancestor, as evident by the domination of the blue region in each of the bars. In other words, the epithelial population is mainly responsible for the proliferation of other cell types. The other phenotypic populations contribute approximately an equal amount to the total population as well as to the individual population per each phenotype, albeit to a much lesser degree than the epithelial population. Figure 4.2 enables the observation of the ancestry of cells with respect to one specific phenotype while Figure 4.3 provides us more insight into the ancestry of these cells with respect to the total population.

4.2.3 Different types of noise generate distinct energy landscapes of the heterogeneous population

In the case of Type 1 noise where the gene expression fluctuations are correlated in each daughter cell, we do not obtain any transition from one cell state to another in the EMT direction. Since a more faithful multiscale model of the EMT process should involve cell transitions in both directions, we will exclude this version of Type 1 noise for the remaining of our analysis.

We further examine the effects of the two types of noise on a collection of cells that constantly undergoes divisions and transitions between states, by studying the results acquired from performing Principal Component Analysis (PCA) on the data derived from the gene expression levels of all the cells. Figure 4.5 is a visualization of the data obtained from two principal components for each type of noise at the end of the second, fifth, and seventh

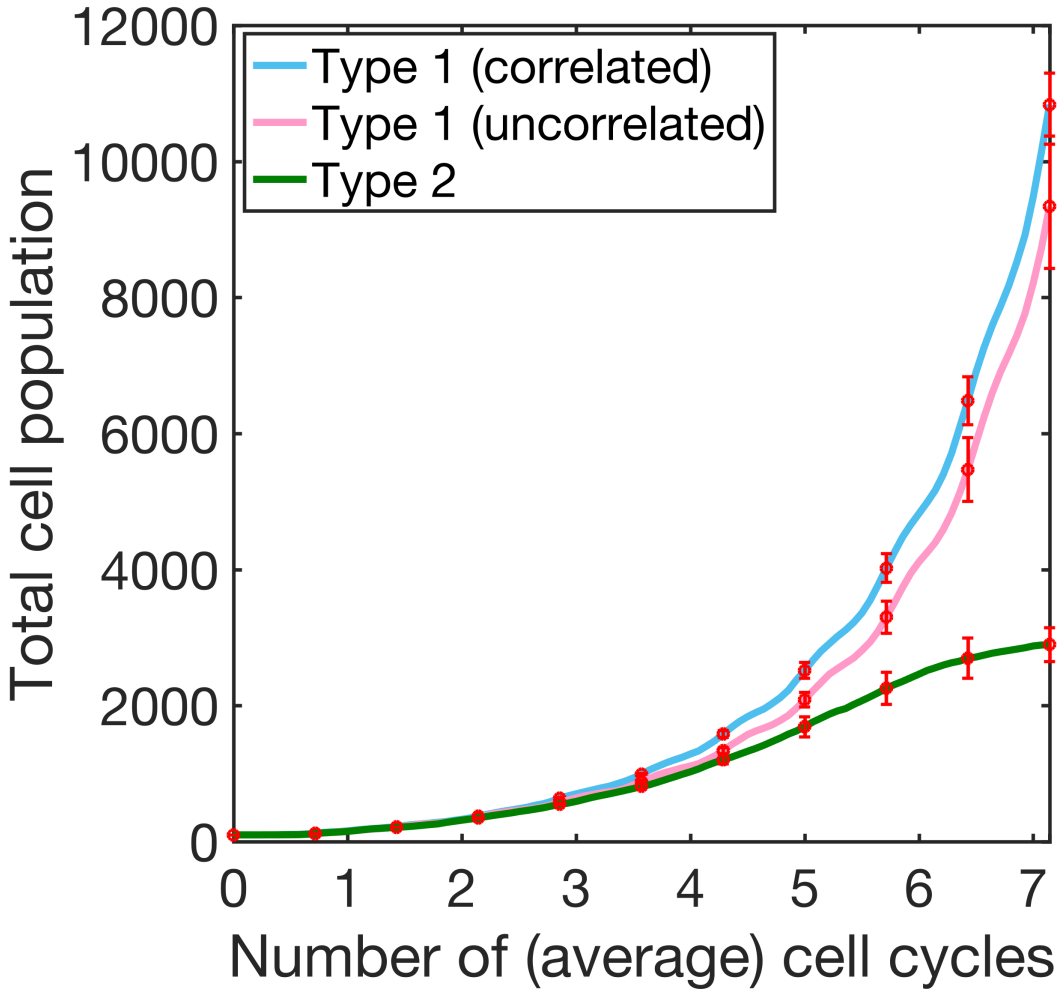

Figure 4.4: The growth of the cell population produced by each type of noise over a period of seven cell cycles. The results are obtained from two simulations per noise type and are plotted using the mean value of these two simulations. The red error bars depict the standard deviation values obtained from two simulations for each type of noise.

division cycles. We are able to identify the structure of a heterogeneous population of cells belonging to different phenotypes in terms of their gene expression levels and follow the evolution of this structure over many division cycles.

From Figure 4.5A, we observe that under the influence of Type 1 noise (uncorrelated version), the many individual cells group themselves into four distinct clusters according to their appropriate phenotypes. The four different phenotypic groups of cells are distinguished by the color scheme mentioned above: blue for E state, green for I_1 state, yellow for I_2 state, and red for M state. As expected, cells that are identified as I_1 cells form the biggest cluster in the PCA plot due to the facilitation of their proliferation by this type of noise. Throughout the seven cell cycles, the four clusters maintain the same approximate distance from one another, with the M cell cluster having a prominent separation from the other three clusters, especially at the end of the seventh cycle. This implies that the M phenotype is relatively more stable compared to the others and there is more difficulty for transitions between M state and other states to take place. It is important not to associate the results of Figure 4.5 with the results of Figure 4.3. Figure 4.3 provides information on which portion of a cell population of a particular phenotype can trace their lineage back to one of the phenotypes included in the initial population, but does not take into account the transitions between states from one cycle to the next.

On the other hand, Type 2 noise does not separate the different phenotypes into distinctive clusters but separates the cell population into two distinct groups characterized by the two different trajectories that they follow for their transitions between one state to another. We observe from Figure 4.5B that at the end of the second cycle, two courses of cell transitions begin to take shape. By the end of the fifth cycle, these two trajectories are clearly defined. The upper trajectory is defined by four distinct steady states in the transition between epithelial and mesenchymal states. Cells that belong to the upper trajectory move through four different states during their transition from E state to M state. Meanwhile, cells from the lower trajectory skip the I_1 state during their transition. Interestingly, the two trajectories

start to merge once the cells in the upper trajectory pass the I_1 state, signifying that the paths to cell transition only differ at the beginning.

In order to have a clearer insight of how the different behaviors of the model under two types of noise translate to distinct differentiation propensities of cells when subjected to various stochastic effects, we implemented an algorithm developed by Chunhe Li [63, 64] and the MATLAB code written by Suoqin Jin to draw a 3D global potential landscape of the data derived from PCA. We present a 3D landscape corresponding to each plot of the gene expression levels in two principal components featured in Figure 4.6 and for each type of noise. When we apply Type 1 noise to our multiscale model, the landscape topography is characterized by two wide potential energy wells where E and I_1 states are, and two narrower wells for the other two states by the end of the second cell cycle, Figure 4.6A. The narrowness of the I_2 well and especially of the M well imply that these two steady states are more stable than the other states in the transition. In addition, the barrier height between E and I_1 states is low, with the barrier height between I_1 and I_2 state being significantly higher, signifying that transitions between E and I_1 states are relatively more likely than transitions between I_1 and I_2 states. Consistent with the 2D visualization of the gene expression levels in Figure 4.5A, the basin of attraction for the M state is completely removed from the other states, implying that transitions from other states to the M state are virtually impossible. In terms of the differentiation propensities of the two intermediate states, both I_1 and I_2 cells are more likely to differentiate to E state. By the end of the fifth cell division cycle, we observe that all the barrier heights between all the states have been significantly lowered, with all the local basins of attractions being equally large. Although the barrier height between I_2 and M state is still relatively high compared to the barriers between the other states, transitions between these two states are now possible. The barrier height between E and I_1 states is slightly lower than than that between I_1 and I_2 states. We observe that I_1 cells still favor transition in the MET direction while I_2 cells are still more likely to convert into I_1 cells than to differentiate into M cells. By the end of the seventh cycle, the M potential energy

well is now narrow again, making transitions from and to this state unlikely.

With Type 2 noise, the landscape topography remains mostly unchanged throughout the seven cycles. At the end of the second, fifth, and seventh cycles, we observe three potential attractors, with the E well being most shallow and the I_2 and M wells being more stable, Figure 4.6B. There is no I_1 well, indicating that I_1 state is possibly only a transient state in this multiscale model with this type of noise. The lack of I_1 attractor helps elucidate Figure 4.5B, where there are two courses of cell transitions. There are two possible trajectories for a cell's transition, with one involving the transient state I_1. The topography of this landscape is also consistent with the shrinking percentage of the population that is I_1 cells over time and signifies that the population of stem cells makes up a very small portion of the cell population. In addition, the positions of the E and I_2 wells are such that epithelial cells cascade down into the I_2 basin of attraction, with a significantly lower barrier height between E and I_2 than that between I_2 and M wells. This implies that epithelial cells are more likely to spontaneously convert to cells in the intermediate state I_2 and settle there. It should be noted that some of the E cells will become I_1 cells momentarily during that conversion. In short, there is a less likelihood of these epithelial cells to spontaneously undergo full EMT. Generally, with Type 1 noise, the energy that a cell must acquire to move to a different steady state is significantly higher than that required for cellular state transitions under Type 2 noise. These statistics corroborate and explain the results presented in Figures 4.2 and 4.3 that the rate of cell transitions between different states is generally lower in the case of Type 1 noise. This apparent disparity in the properties of the two noises implies that there are many more aspects of the multiscale model we need to consider before achieving a faithful representation of the EMT process, especially how the different types of noise can be combined and modulated.

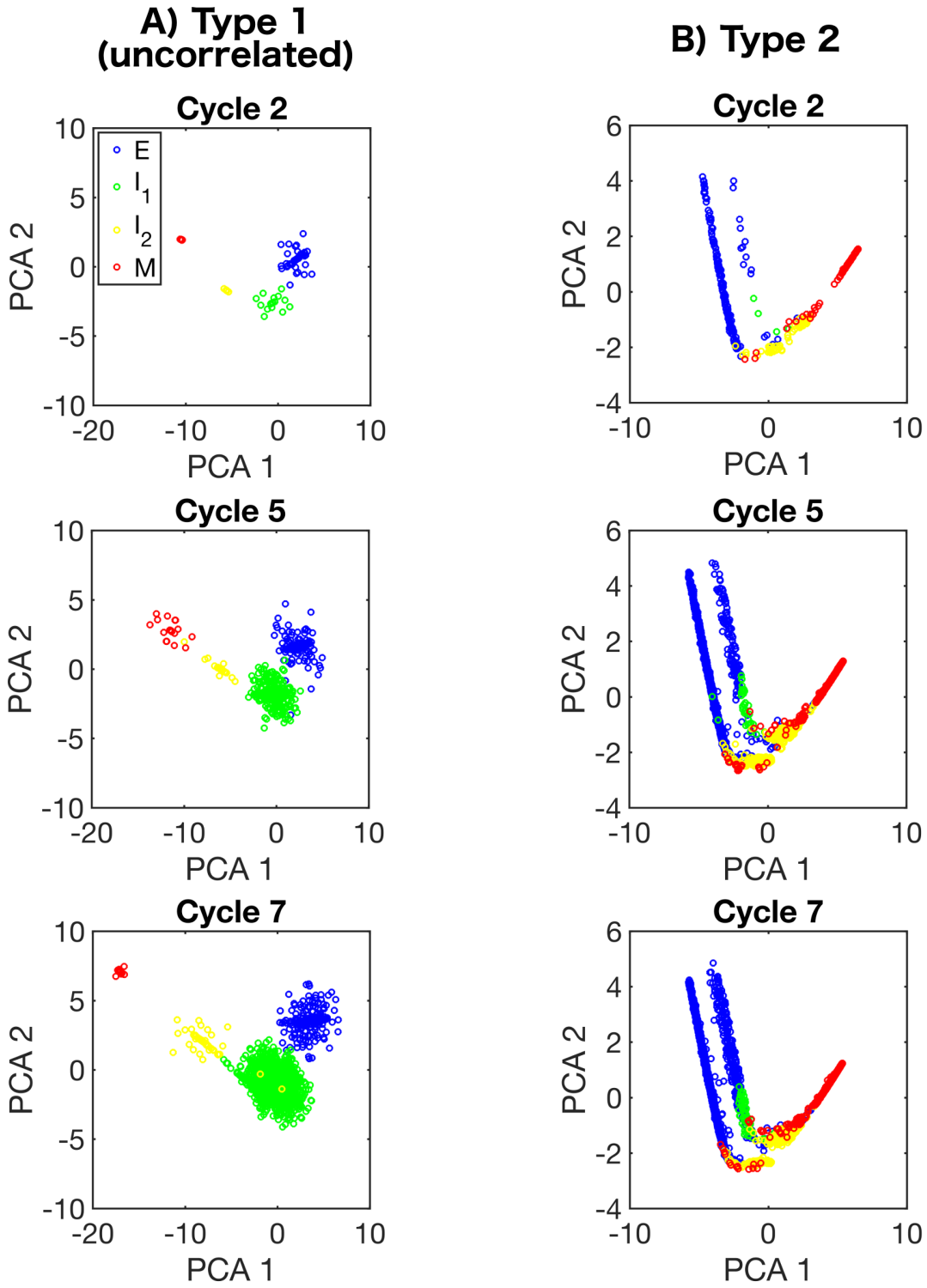

Figure 4.5: The 2D visualization of the heterogeneous population at the end of the second, fifth, and seventh cycles obtained from one simulation of cellular activities for **A)** Type 1 (uncorrelated) noise and **B)** Type 2 noise with respect to the first and second principal components (PC1 and PC2).

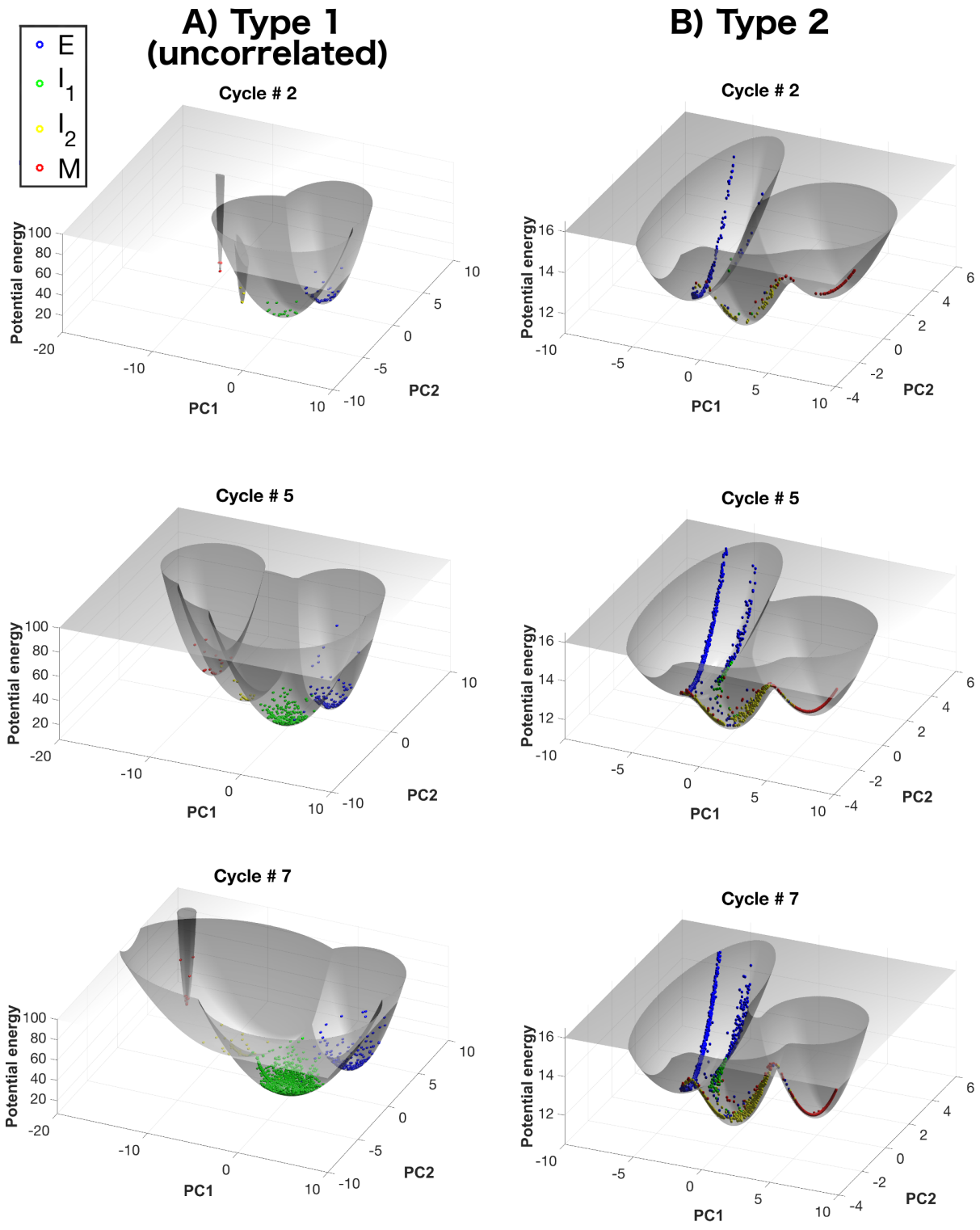

Figure 4.6: The 3D potential landscapes of the heterogeneous population at the end of the second, fifth, and seventh cycles obtained from one simulation of cellular activities for **A)** Type 1 (uncorrelated) noise and **B)** Type 2 noise with respect to the first and second principal components (PC1 and PC2). Each well in every subplot is a local basin of attraction.

4.3 Discussion

Previous research studies the regulation mechanisms of the EMT process using a framework based on the regulatory network. This framework involves modeling the core network using a system of ODEs with feedback control to identify important agents that exert the most control on the cellular state transitions and essentially are responsible for initiating/ suppressing EMT [40, 47, 48, 76, 100, 115]. The limitations imposed on this framework is that the cellular activities defined by changes in gene expressions which further represent changes in the cell state, are confined within the context of one individual cell. Essentially, the framework focuses only on the transition of one cell between one cell state and the next. Our recent efforts to approach the EMT process from the population dynamics angle, by monitoring the evolution of the different populations in the EMT spectrum over time, result in the demonstration of a correlation between the number of intermediate states and the noise-attenuating ability of the overall cell population. More specifically, increasing the number of intermediate states in the EMT spectrum helps reduce the fluctuations in the population [97]. However, our model excluded the individual cellular dynamics prompted by the interactions between different EMT promoting and suppressing agents.

We attempted to extend the modeling of the EMT process beyond the scope of one cell to that of a heterogeneous population of cells while incorporating the gene regulatory network into our multiscale model. In our model, each cell was equipped with the ability to divide and proliferate, or differentiate, as well as subjected to cellular state changes triggered by stochastic effects. We attributed finite division capacity to non-stem cells and designated a population of one of the intermediate states, I_1, to be the stem cell population. We explored the relationship between stem cells with the whole population, as well as elucidated its role in regenerating the depleted non-stem cell population from cell death. In order for stem cells to replenish the other cellular populations, some of the stem cells must undergo differentiation, hence, the introduction of noise to induce cellular state transitions. We experimented

with two types of noise: Type 1 noise, which involves subjecting the mother cell to gene fluctuations upon division, and Type 2 noise, which is characterized by parametric noise at cell division.

We found that Type 2 noise encouraged more cell state transitions in both EMT and MET direction than Type 1 noise. Type 1 noise is further split into two subcategories. When all the fluctuations in all the gene expressions of the mother cell upon division are uncorrelated in the daughter cell, there are transitions in both EMT and MET direction, albeit relatively fewer cells change fate towards a more mesenchymal direction. When the fluctuations are correlated so that the gene expressions all increase or decrease in a daughter cell, there exist transitions in the MET direction but no transitions towards a more mesenchymal potential. In addition, Type 2 noise endows the stem cell population, i.e I_1 population, with higher differentiation and asymmetric division potentials, leading to a quickly diminishing stem-cell proportion in the population, and consequently a significantly lower total population than that obtained from Type 1 noise. On the contrary, for Type 1 noise, a very small percentage of cells in each phenotypic group can trace their origins back to a stem cell in the initial population. Next, we showed that the 3D landscape topography with respect to two principal components of our model under Type 1 noise demonstrated the organization of the population into four distinct clusters according to four different phenotypes. Throughout the seven cycles, M phenotype remains the most stable with the least potential for cells to transition to and from. Meanwhile, the I_1 and I_2 phenotypes become less stable over time, resulting in more cell state transitions between I_1 and I_2 states as well as from I_1 to E states by the end of the seventh division cycle.

On the other hand, when examining the 3D landscape generated by Type 2 noise, we observe that the cell population arrange themselves according to two possible paths they can commit to: one without the I_1 state, and the other with a transient I_1 state. The potential energy well for the I_1 state is non existent in the second path. Coupled with the low energy barrier between the E and the I_2 wells, these results explain the diminishing percentage of

stem cells in the population over time. In general, when subjected to Type 1 noise, cells need to overcome significantly higher energy barriers to initiate a state transition than when under the influence of Type 2 noise. As a consequence, there is a much higher rate of cell transitions between all the states in both EMT and MET directions with Type 2 noise than with Type 1 noise.

The consistently different results arisen from introducing different types of noise into our multiscale model illustrate the need to experiment with other noises, such as additive or multiplicative noises on gene expressions to explore the strategic significance of different types of noise in shaping the structure and distribution of a heterogeneous cellular population, as well as in endowing different differentiation potentials to the stem cell population. To reproduce a more faithful multiscale model of the EMT process, an inclusion of various types of noise as well as experimental data in our multiscale model is necessary.

In short, our work attempts to bridge the gap between discrete analysis and population dynamics modeling by incorporating the EMT core regulatory network into our previous population model. In essence, our modeling work offers a framework for studying the regulation mechanism of cell proliferation through monitoring the changes in gene expression levels prompted by interactions between various EMT agents.

Chapter 5

An Integration Factor Method for Stochastic Reaction-Diffusion Systems

5.1 Implicit Integration Factor methods

5.1.1 Construction of General Method

We consider the stiff reaction-diffusion equation with spatial white noise below:

$$\frac{\partial U}{\partial t} = a\frac{\partial^2 U}{\partial x^2} + f(U) + g(U)\frac{\partial^2 W}{\partial x \partial t} \tag{5.1}$$

where $a\partial^2 U/\partial x^2$ is the diffusion term and a is a nonnegative constant, $f(U)$ is the reaction term, and $g(U)(\partial^2 W/\partial x\partial t)$ is the noise term of two possible forms: $g(U)(\partial^2 W/\partial x\partial t) = \sigma(\partial^2 W/\partial x\partial t)$ for additive noise or $g(U)(\partial^2 W/\partial x\partial t) = \sigma U(\partial^2 W/\partial x\partial t)$ for multiplicative noise. Here, σ is a constant to describe the level of noise. Also, $\partial^2 W/\partial x\partial t$ denotes the mixed second-order derivative of the Brownian sheet. A one-dimensional Brownian sheet is

a 2-parameter, centered Gaussian process $B = B(s,t); s,t > 0$ whose covariance is given by:

$$\mathbb{E}(B(s,t)B(s',t')) = \min(s,s') \times \min(t,t'), \forall s,s',t,t' \geq 0. \tag{5.2}$$

Before discussing the derivation of the numerical methods to solve Eq.(5.1), we briefly review the Implicit Integration Factor methods (IIF) discussed in [83], which is crucial to the construction of the IIF methods for a stochastic system. Using the semi-discretized form $dU/dt = aU + f(U)$ that is obtained after the discretization of the diffusion operator in space, we multiply both sides of the equation by the integrating factor e^{-at} and integrating the equation over one time step from t_n to $t_{n+1} = t_n + \triangle t$ to get

$$U(t_{n+1}) = U(t_n)e^{a\triangle t} + e^{a\triangle t}\int_0^{\triangle t} e^{-a\tau} f(U(t_n+\tau))d\tau \tag{5.3}$$

Using appropriate approximation of the integrands in $\int_0^{\triangle t} e^{-a\tau} f(U(t_n+\tau))d\tau$ one derives r^{th}-order IIF scheme [83]:

$$U_{n+1} = e^{a\triangle t}U_n + \triangle t\left(\alpha_{n+1}f(U_{n+1}) + \sum_{i=0}^{r-2}\alpha_{n-i}f(U_{n-1})\right), \tag{5.4}$$

with α_{n+1},α_n, α_{n-1},...,α_{n-r+2} defined as

$$\alpha_{n-i} = \frac{e^{(i+1)a\triangle t}}{\triangle t}\int_0^{\triangle t}\prod_{j=-1,j\neq i}^{r-2}\frac{\tau + j\triangle t}{(j-i)\triangle t}d\tau, \qquad -1 \leq i \leq r-2. \tag{5.5}$$

Similarly for the stochastic reaction-diffusion systems, we first discretize the space using m points with the spatial interval $\triangle x$. Let U_t be a vector whose i^{th}-entry is the value of the solution to Eq.(5.1) at the i^{th} spatial point. A second-order central difference approximation of $\partial^2 U/\partial x^2$ in Eq.(5.1) with periodic boundary condition $U(x_0, t) = U(x_f, t)$ on the SPDE as in [25], where x_0 and x_f indicate the endpoints of the spatial interval, leads to

$$dU_t = aMU_t dt + f(U_t)dt + g(U_t)\frac{dW_t}{\sqrt{\triangle x}} \tag{5.6}$$

where

$$M = \frac{1}{(\triangle x)^2}\begin{pmatrix} -2 & 1 & 0 & 0 & \cdots & 1 \\ 1 & -2 & 1 & 0 & \cdots & 0 \\ 0 & 1 & -2 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 & -2 & 1 \\ 1 & 0 & \cdots & 0 & 1 & -2 \end{pmatrix}. \tag{5.7}$$

Let $\bar{g}(U_t) = g(U_t)/\sqrt{\triangle x}$ and multiply both sides of this Eq.(5.6) by the integrating factor e^{-aMs}. We then have

$$e^{-aMs}dU_{t_n+s} = aMe^{-aMs}U_{t_n+s}ds + e^{-aMs}f(U_{t_n+s})ds + e^{-aMs}\bar{g}(U_{t_n+s})dW_s \tag{5.8}$$

$$e^{-aMs}dU_{t_n+s} - aMe^{-aMs}U_{t_n+s}ds = e^{-aMs}f(U_{t_n+s})ds + e^{-aMs}\bar{g}(U_{t_n+s})dW_s \tag{5.9}$$

$$\int_0^{\triangle t}[e^{-aMs}dU_{t_n+s} - aMe^{-aMs}U_{t_n+s}ds] = \int_0^{\triangle t} e^{-aMs}f(U_{t_n+s})ds + \int_0^{\triangle t} e^{-aMs}\bar{g}(U_{t_n+s})dW_s \quad (5.10)$$

$$\int_0^{\triangle t} d(e^{-aMs}U_{t_n+s}) = \int_0^{\triangle t} e^{-aMs}f(U_{t_n+s})ds + \int_0^{\triangle t} e^{-aMs}\bar{g}(U_{t_n+s})dW_s \quad (5.11)$$

Taking the integral of the left side gives

$$e^{-aM\triangle t}U_{t_{n+1}} - U_{t_n} = \int_0^{\triangle t} e^{-aMs}f(U_{t_n+s})ds + \int_0^{\triangle t} e^{-aMs}\bar{g}(U_{t_n+s})dW_s \quad (5.12)$$

Letting $\triangle t = t_{n+1} - t_n$ and with some more simplification, the equation above becomes

$$U_{t_{n+1}} = e^{aM\triangle t}\left(U_{t_n} + \int_0^{\triangle t} e^{-aMs}f(U_{t_n+s})ds + \int_0^{\triangle t} e^{-aMs}\bar{g}(U_{t_n+s})dW_s\right) \quad (5.13)$$

All we have left is evaluating the right side of Eq.(5.13). Observe the following numerical approximation of the noise part of Eq.(5.13) [38]

$$\int_0^{\triangle t} e^{-aMs}\bar{g}(U_{t_n+s})dW_s = \bar{g}(U_{t_n})(W_{t_{n+1}} - W_{t_n}) \quad (5.14)$$

To approximate the deterministic part of Eq.(5.13), i.e $e^{aM\triangle t}\left(U_{t_n} + \int_0^{\triangle t} e^{-aMs}f(U_{t_n+s})ds\right)$, we apply the IIF strategy using Eqs.(5.4) and (5.5). Coupling this evaluation with Eq.(5.14),

Eq.(5.13) becomes

$$U_{n+1} = e^{aM\triangle t}U_n + \triangle t\left(\alpha_{n+1}f(U_{n+1}) + \sum_{i=0}^{r-2}\alpha_{n-i}f(U_{n-1})\right) + e^{aM\triangle t}\bar{g}(U_n)\triangle W_n \quad (5.15)$$

where $\triangle W_n = W_{t_{n+1}} - W_{t_n}$, $U_{t_n} = U_n$, and α_{n+1},α_n, α_{n-1},...,α_{n-r+2} are as described in Eq.(5.5).

Let us denote $\bar{\xi_n}$ to be a standard normally-distributed random vector and n to be the indices of the temporal discretization points. We apply the standard Maruyama method to the noise term along with the first order IIF method, denoted as IIF1, or the second order IIF method, denoted as IIF2, to obtain

IIF1-Maruyama method

$$U_{n+1} = e^{aM\triangle t}U_n + \triangle t f(U_{n+1}) + e^{aM\triangle t}g(U_n)\frac{\sqrt{\triangle t}\bar{\xi_n}}{\sqrt{\triangle x}} \quad (5.16)$$

IIF2-Maruyama method

$$U_{n+1} = e^{aM\triangle t}\left(U_n + \frac{1}{2}\triangle t f(U_n)\right) + \frac{1}{2}\triangle t f(U_{n+1}) + e^{aM\triangle t}g(U_n)\frac{\sqrt{\triangle t}\bar{\xi_n}}{\sqrt{\triangle x}} \quad (5.17)$$

When the stochastic integral in Eq.(5.13) is approximated explicitly as in Eq.(5.14), the strong order of convergence of the overall scheme is dominated by the root mean-square order of the increments $\triangle W_n$, which is one-half [54]. For this reason, the strong order of convergence for both of our methods will be consistent with those of most other methods with the same approximation of the stochastic term, i.e the Euler Maruyama method. Let

us illustrate this through approximating an SODE of the similar form

$$dU_t = -aU_t dt + bU_t dt + g(U_t)dW_t \tag{5.18}$$

using IIF1 and IIF2 methods respectively. By using the standard Maruyama approximation on the noise term, we obtain

IIF1-Maruyama method:

$$U_{n+1} = e^{-a\triangle t}U_n + b\triangle t U_{n+1} + g(U_n)e^{-a\triangle t}\triangle W_n \tag{5.19}$$

IIF2-Maruyama method:

$$U_{n+1} = e^{-a\triangle t}U_n + \frac{1}{2}e^{-a\triangle t}b\triangle t U_n + \frac{1}{2}b\triangle t U_{n+1} + g(U_n)e^{-a\triangle t}\triangle W_n \tag{5.20}$$

When the noise is additive, i.e.

$$g(U_t) = \sigma, \tag{5.21}$$

the strong order of convergence for both methods is one, which is consistent with the strong order of convergence of the Euler Maruyama method [54].

When the noise is multiplicative, i.e.

$$g(U_t) = \sigma U_t \tag{5.22}$$

both methods share the same order of convergence of one-half with the Euler Maruyama method [54].

5.1.2 Stability Analysis

Additive noise

We first analyze the numerical stability of the IIF1-Maruyama and IIF2-Maruyama schemes in Eqs.(5.19) and (5.20) when the stochastic differential equation has additive noise Eq.(5.21).

Following a previous study [54], we define $U_{n_t}^{\delta}$ to be a time discrete approximation of the solution $U(t)$ with maximum step size $\delta > 0$ starting at time t_0 at U_0^{δ} and $\bar{U}_{n_t}^{\delta}$ to be the corresponding approximation starting at $\bar{U}_0^{\delta}$. Then $U_{n_t}^{\delta}$ is *asymptotically numerically stable* for a given stochastic differential equation if for any finite interval $[t_0, T]$ there exists a positive constant $\triangle_a$ such that for each $\epsilon > 0$ and $\delta \in (0, \triangle_a)$ [54]:

$$\lim_{|U_0^{\delta} - \bar{U}_0^{\delta}| \to 0} \sup_{t_0 \le t \le T} P\Big(|U_{n_t}^{\delta} - \bar{U}_{n_t}^{\delta}| \ge \epsilon \Big) = 0 \tag{5.23}$$

and

$$\lim_{|U_0^{\delta} - \bar{U}_0^{\delta}| \to 0} \lim_{T \to \infty} P\Big(\sup_{t_0 \le t \le T} |U_{n_t}^{\delta} - \bar{U}_{n_t}^{\delta}| \ge \epsilon \Big) = 0 \tag{5.24}$$

with $P(A)$ indicating the probability that event A occurs. We can analyze the asymptotic stability of a numerical stochastic scheme as we do for the A-stability of deterministic differential equations by studying the stability of the following class of complex-valued linear test equations [54]:

$$dU_t = \lambda U_t dt + dW_t \tag{5.25}$$

where λ is a complex number with $\mathbb{R}(\lambda) < 0$ and W is a real-valued standard Wiener process.

Suppose that a numerical scheme with equidistant step size $\triangle t \equiv \delta$ applied to test equation (5.25) with $\mathbb{R}(\lambda) < 0$ can be written recursively as:

$$U_{n+1}^{\triangle t} = G(\lambda \triangle t) U_n^{\triangle t} + Z_n^{\triangle t} \tag{5.26}$$

where G is a mapping of complex plane $\mathbb{C}$ into itself and $Z_n^{\triangle t}$ are random variables that do not depend on $U_n^{\triangle t}$ for $n = 0, 1, 2, ...$, then the set of complex values $\lambda \triangle t$ satisfying

$$\mathbb{R}(\lambda) < 0 \text{ and } |G(\lambda \triangle t)| < 1 \tag{5.27}$$

is the region of absolute stability of that scheme [54].

Our methods Eqs.(5.19) and (5.20) when applied to the linear test equation Eq.(5.25) are reduced to:

$$U_{n+1}^{\triangle t} = e^{\lambda \triangle t} U_n^{\triangle t} + \sigma e^{\lambda \triangle t} \triangle W_n \tag{5.28}$$

Since $|G(\lambda \triangle t)| = |e^{\lambda \triangle t}| < 1$ for any arbitrarily large value $\triangle t > 0$ given $\mathbb{R}(\lambda) < 0$, we can claim that both the IIF1-Maruyama and IIF2-Maruyama methods are absolutely stable when noise is additive.

Multiplicative Noise

When the noise is multiplicative Eq.(5.22), we analyze the stability of each method using mean-square stability analysis [54]. A method is mean-square stable if $lim_{n \to \infty} \mathbb{E}(|U_n|^2) = 0$ where $\mathbb{E}(.)$ denotes the expected value. To apply this technique to evaluating the stability region of both IIF-Maruyama methods aforementioned, we note that we can rewrite each

method in the form:

$$U_{n+1} = h(a, b, \sigma, \triangle t, \triangle W_n)U_n \tag{5.29}$$

Squaring and then taking expectation of both sides of Eq.(5.29) coupled with the fact that W_t is a standard Wiener process whose increment $W(t) - W(s)$ is normally-distributed with mean 0 and variance $t - s$, we obtain

$$\mathbb{E}|U_{n+1}|^2 = H(a, b, \sigma, \triangle t)\mathbb{E}|U_n|^2 \tag{5.30}$$

where $H(a, b, \sigma, \triangle t) = \mathbb{E}(h(a, b, \sigma, \triangle t, \triangle W_n))^2$.

Eq.(5.30) demonstrates that $lim_{n\to\infty}\mathbb{E}(|U_n|^2) = 0$, i.e, the numerical method is mean-square stable if and only if $H(a, b, \sigma, \triangle t) < 1$ [37].

For the IIF1-Maruyama method, the mean-square stability condition becomes

$$e^{-2a\triangle t}(1 + \sigma^2\triangle t) - (1 - b\triangle t)^2 < 0 \tag{5.31}$$

For the IIF2-Maruyama method, the mean-square stability condition becomes

$$(2 + b\triangle t)^2 - (2 - b\triangle t)^2 e^{2a\triangle t} + 4\sigma^2\triangle t < 0 \tag{5.32}$$

We plot the stability regions of both IIF-Maruyama schemes on a plane whose axes are $a\triangle t$ and $b\triangle t$ in Figure 5.1 and vary the value of $\sigma^2\triangle t$. Note that the stability region in Figure 5.1 for each method is the region under the respective colored curve. The desired absolute

stability region is the region where the diffusion and reaction coefficients are negative. In terms of Eq.(5.18), this region is described as $\{(a, b) : a > 0 \text{ and } b < 0\}$. In Figure 5.1A when there is no noise term, both methods are unconditionally stable with respect to this absolute stability region which is the inside of the square with dashed border. From Figure 5.1B and C, we observe that as the value of $\sigma^2 \triangle t$ increases, the stability region of the IIF2-Maruyama method shrinks at a faster rate than the stability region of the IIF1-Maruyama method, resulting in the IIF1-Maruyama method having a larger stability region when the noise term is large enough. As a result, the IIF1-Maruyama method has a more desirable stability than the IIF2-Maruyama method in the case of more dominant noise.

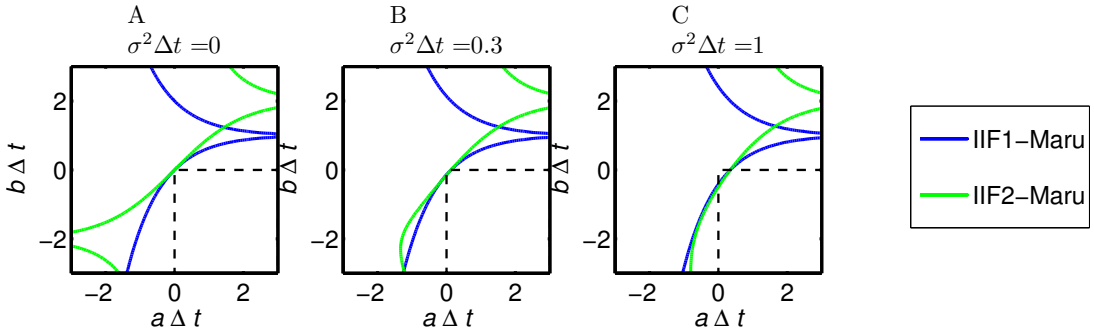

Figure 5.1: The stability regions of both IIF-Maruyama methods described in Eqs.(5.16) and (5.17) for multiplicative noise. The stability region lies below the corresponding colored curve. The desired absolute stability region is the region inside the square with dashed-border.

Comparison with other methods in the case of Multiplicative Noise

For the purpose of stability-region comparison, we present three other methods used to solve Eq.(5.18) and their constructions: The Euler Maruyama method [54] when it is applied to Eq.(5.18) takes the form

Euler Maruyama method

$$U_{n+1} = U_n - a\triangle t U_n + b\triangle t U_n + g(U_n)\triangle W_n \tag{5.33}$$

The order of accuracy for this method is 1/2 [54] and the mean-square stability analysis when noise is multiplicative, i.e Eq.(5.22) gives the following stability condition:

$$-2a\triangle t + (a\triangle t)^2 - 2(a\triangle t)(b\triangle t) + 2b\triangle t + (b\triangle t)^2 + \sigma^2\triangle t < 0 \tag{5.34}$$

The next method is designed in a similar fashion to the construction of the IIF-Maruyama methods with a modification on the approximation of the deterministic integral term of Eq.(5.13). Direct application of the exponential time differencing method of order 2 on this term leads to

ETD2-Maruyama method

$$U_{n+1} = \left(e^{-a\triangle t} + \frac{b}{a}(1 - e^{-a\triangle t})\right)U_n + e^{-a\triangle t} g(U_n)\triangle W_n \tag{5.35}$$

Since the stochastic integral term is approximated explicitly as in the IIF-Maruyama methods, the order of accuracy for the overall scheme is 1/2. Mean-square stability analysis gives the following stability condition for the above method:

$$e^{-2a\triangle t} + 2(e^{-a\triangle t} - e^{-2a\triangle t})\frac{b\triangle t}{a\triangle t} + \left((1 - e^{-a\triangle t})\frac{b\triangle t}{a\triangle t}\right)^2 + \sigma^2\triangle t e^{-2a\triangle t} - 1 < 0 \tag{5.36}$$

The last scheme mentioned here is constructed similarly to the Euler Maruyama method with the exception of the deterministic term being approximated using second-order Runge Kutta method.

RK2-Maruyama method

$$U_{n+1} = \left(1 + (b-a)\triangle t + \frac{1}{2}(b-a)^2\triangle t^2\right)U_n + g(U_n)\triangle W_n \tag{5.37}$$

The construction of the RK2-Maruyama scheme also exploits the explicit approximation of the stochastic term as that of the Euler Maruyama scheme, so the order of accuracy for the RK2-Maruyama scheme is still 1/2. The scheme's mean-square stability condition is:

$$2((a\triangle t)^2+(b\triangle t)^2)+2(b\triangle t-a\triangle t)-4(ab)(b\triangle t)+(b\triangle t)^3-3(b\triangle t)^2(a\triangle t)+3(b\triangle t)(a\triangle t)^2-(a\triangle t)^3$$

$$+\frac{1}{4}(b\triangle t)^4-(b\triangle t)^3(a\triangle t)+\frac{3}{2}(a\triangle t)^2(b\triangle t)^2-(b\triangle t)(a\triangle t)^3+\frac{1}{4}(a\triangle t)^4+\sigma^2\triangle t<0 \tag{5.38}$$

To illustrate the performance of the IIF-Maruyama schemes in terms of stability analysis in comparison with the above methods, we plot all the stability regions of each method for different values of $\sigma^2\triangle t$ on a plane whose axes are $a\triangle t$ and $b\triangle t$ (Figure 5.2). In the same figure, the region where unconditional stability is achieved for an ideal method is the region inside the box with dashed boundary. Figure 5.3 is the enlarged version of Figure 5.2 so we can observe better the changes in the absolute stability region for each method at different values of $\sigma^2\triangle t$.

In Figure 5.2A, i.e when there is no noise term, only the IIF-Maruyama methods are unconditionally stable, which is consistent with the stability of the deterministic IIF methods. For very positive values of the diffusion term and very negative values of the reaction terms (in terms of Eq.(5.18), this means that both $a < 0$ and $b < 0$), the Euler Maruyama, RK2-Maruyama, and ETD2-Maruyama methods achieve better stability than the IIF-Maruyama methods, as seen in the bottom left corner of each subplot of Figure 5.2A. However, the

overall size of the absolute stability regions of the IIF-Maruyama methods is still larger than those of the other methods.

With the increasing size of the multiplicative noise term, the stability region for each method starts to shrink. More specifically, the stability regions for the Euler Maruyama and RK2-Maruyama methods start to shrink in width along the line $b = -a$ (i.e along the direction in which the diffusion and reaction terms are equal) and become thin strips in Figure 5.2B and C. Both of these regions disappear completely in the next plot, i.e Figure 5.2D when the noise amplitude is high enough.

Meanwhile, the bottom-left corner of the stability region of the ETD2-Maruyama method recedes significantly as the noise amplitude increases, resulting in a greater loss of the absolute stability region than those of the IIF-Maruyama methods. Size comparison of the absolute stability regions from both Figure 5.2B-D and Figure 5.3B-D indicates that the stability region of the ETD2-Maruyama shrinks more than those of the IIF-Maruyama methods as $\sigma^2 \triangle t = 1$ increases from 0 to 1.

From Figure 5.2B-D, we observe that the IIF-Maruyama methods have the greatest region of absolute stability for any positive values of $\sigma^2 \triangle t$. Also, from Figure 5.3B-D, we observe that the stability regions for both IIF-Maruyama methods shrink at a slower rate than those of the other methods. As a result, both methods have the best absolute stability region for large noise amplitude, as demonstrated by Figure 5.2D, when $\sigma^2 \triangle t = 1$. In addition, the absolute stability region of the IIF1-Maruyama method shrinks more slowly than the IIF2-Maruyama, as evidenced by Figure 5.3B-D where the dashed box indicates the ideal absolute stability region. For this reason, the IIF1-Maruyama method achieves the largest absolute stability region for large noise amplitude out of all the methods.

We conclude that at different values of the noise term, the IIF-Maruyama methods outperform the other methods in terms of the region of absolute stability. In particular, at

$\sigma^2 \triangle t = 1$, both the IIF-Maruyama methods have a much greater region of absolute stability than the rest of the methods.

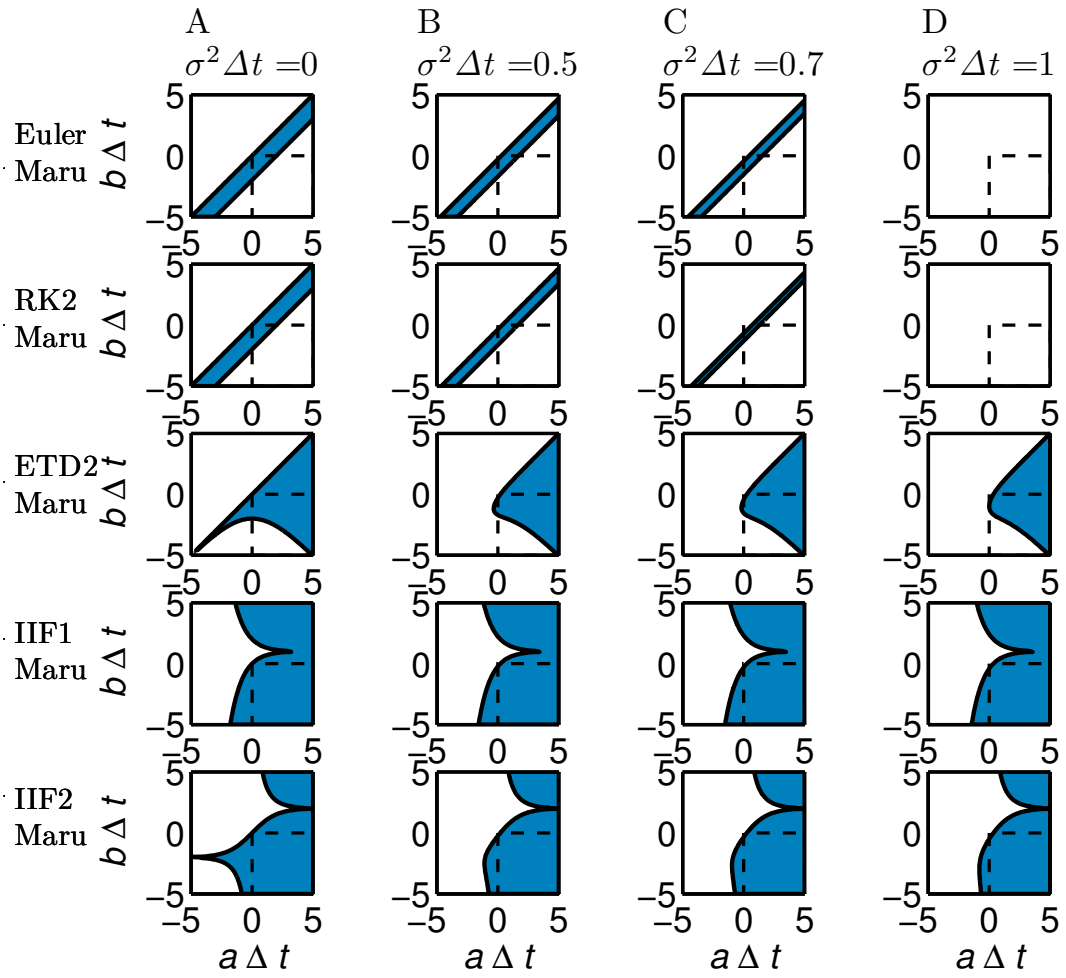

Figure 5.2: Comparison of stability regions of the following methods: IIF1-Maruyama, IIF2-Maruyama, Euler Maruyama, RK2-Maruyama, and ETD2-Maruyama used to solve Eq.(5.18) with multiplicative noise. The stability region for each method is shaded blue. The ideal absolute stability region is the region inside the dashed box.

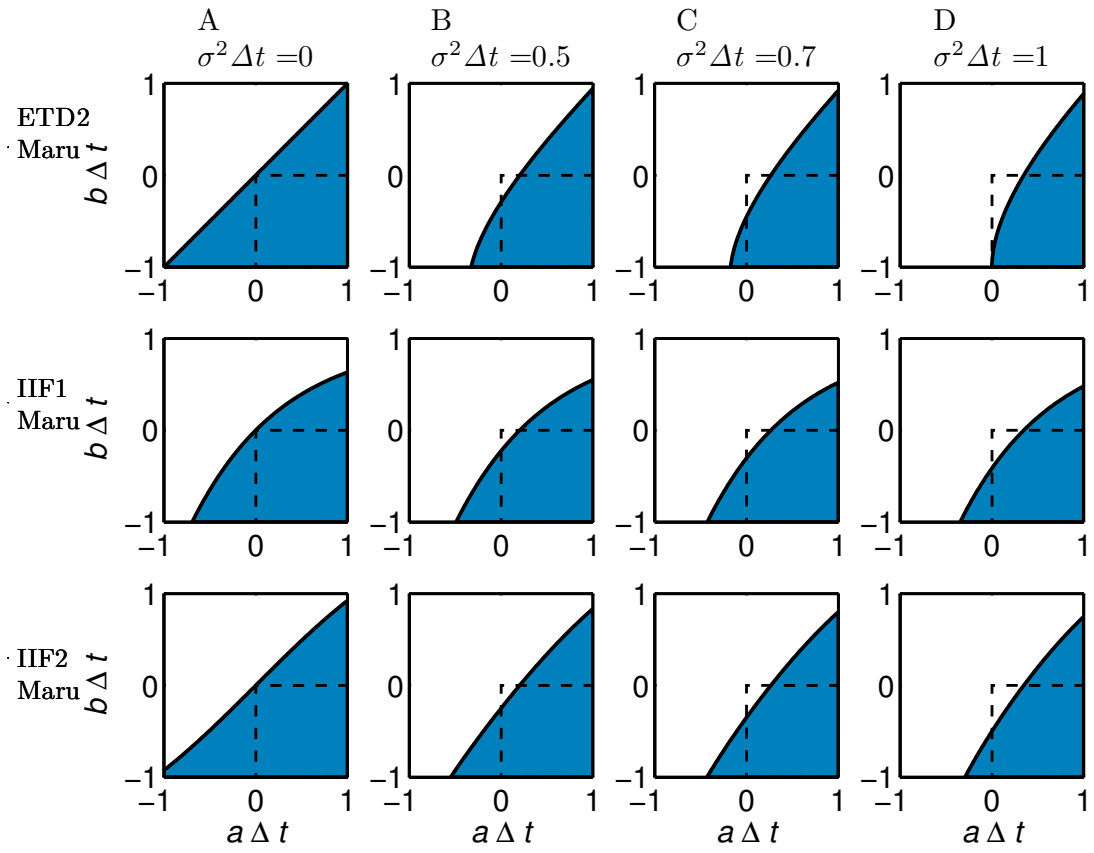

Figure 5.3: Enlarged version of Figure 5.2. In this figure, the stability regions are enlarged for the following methods: IIF1-Maruyama, IIF2-Maruyama, and ETD2-Maruyama. The stability region for each method is shaded blue. The ideal absolute stability region is the region inside the dashed box.

5.2 Numerical Simulations

First, we compare the two IIF-Maruyama methods with the other methods when they are applied to Eq.(5.1) for both cases of additive noise and multiplicative noise. Through choosing different values of a, which corresponds to the size of diffusion, and different values of b, which corresponds to the strength and stiffness of reactions, we evaluate the convergence and stability of IIF-Maruyama methods.

5.2.1 Tests on Stochastic Ordinary Differential Equations

Here, we implement various methods to solve the linear stochastic ODE Eq.(5.18). Comparisons will be made between the two IIF methods, the Euler Maruyama, and the ETD2-Maruyama methods. The comparisons concern the accuracy of these methods in situations where the degree of stiffness is high or the noise amplitude is great. All the simulations are done over 1000 independent paths with a time frame from 0 to 1 unless specified otherwise. Numerical experiments were needed in order to decide on a sufficiently large number of Brownian paths that will yield the desirable orders of convergence. All the results obtained in this sub-section remain consistent for a greater number of Brownian paths. This notion is confirmed when we increase the number of paths from 1000 to 2000 and subsequently 10000.

Additive Noise

Denote $U_{\triangle t}$ to be the solution obtained numerically from using time step $\triangle t$. The order of convergence for additive noise is the value γ such that there exists a constant C where

$$\mathbb{E}|U_{\triangle t} - U_{\triangle t/2}| \leq C \triangle t^{\gamma} \tag{5.39}$$

for $\triangle t$ sufficiently small. The order of convergence for the case, in which the explicit solution (e.g to Eq.(5.18) with additive noise Eq.(5.21)) is unknown, is estimated by

$$\text{Order of Convergence } = \frac{\log\left(\mathbb{E}|U_{\triangle t} - U_{\triangle t/2}| \Big/ \mathbb{E}|U_{\triangle t/2} - U_{\triangle t/4}|\right)}{\log 2}. \tag{5.40}$$

In our simulations, we start with $\triangle t = 2^{-5}$ and decrease $\triangle t$ by half for a total of 6 times. 1000 independent Brownian paths are generated and the final solution $U_{\triangle t}$ on each path for each time step $\triangle t$ is calculated.

Next, we study the accuracy and stablity for both IIF-Maruyama methods and compare them with the Euler Maruyama and ETD2-Maruyama methods in different scenarios, especially, in the case in which the reaction term is dominant and the system becomes stiff.

In Figure 5.4, we plot all the mean errors of the numerical solutions obtained from the IIF-Maruyama, Euler Maruyama, and ETD2-Maruyama methods while using different time steps in the scenario where the reaction term is heavily stiff, i.e when the magnitude of the reaction term is relatively large compared to the magnitudes of the diffusion and noise terms. Here, the mean error is defined to be $\mathbb{E}|U_{\triangle t} - U_{\triangle t/2}|$ where $U_{\triangle t}$ is the numerical solution resulted from using each of the above methods with the time step $\triangle t$. From this figure, we observe that both IIF-Maruyama methods maintain a low mean error as the time step $\triangle t$ increases in size. When $\triangle t$ becomes too large, i.e when $\triangle t = 1/4$, the mean errors of the solutions from using Euler Maruyama and ETD2-Maruyama methods explode out of reasonable bounds. Meanwhile, at the same time step, the mean errors of the numerical solutions resulted from the two IIF-Maruyama methods remain consistently small when larger step size $\triangle t$ is used. This figure demonstrates that the IIF-Maruyama methods are highly effective whenever the reaction-diffusion system has a dominant reaction term.

Figure 5.5 shows the orders of convergence for IIF1-Maruyama, IIF2-Maruyama, Euler

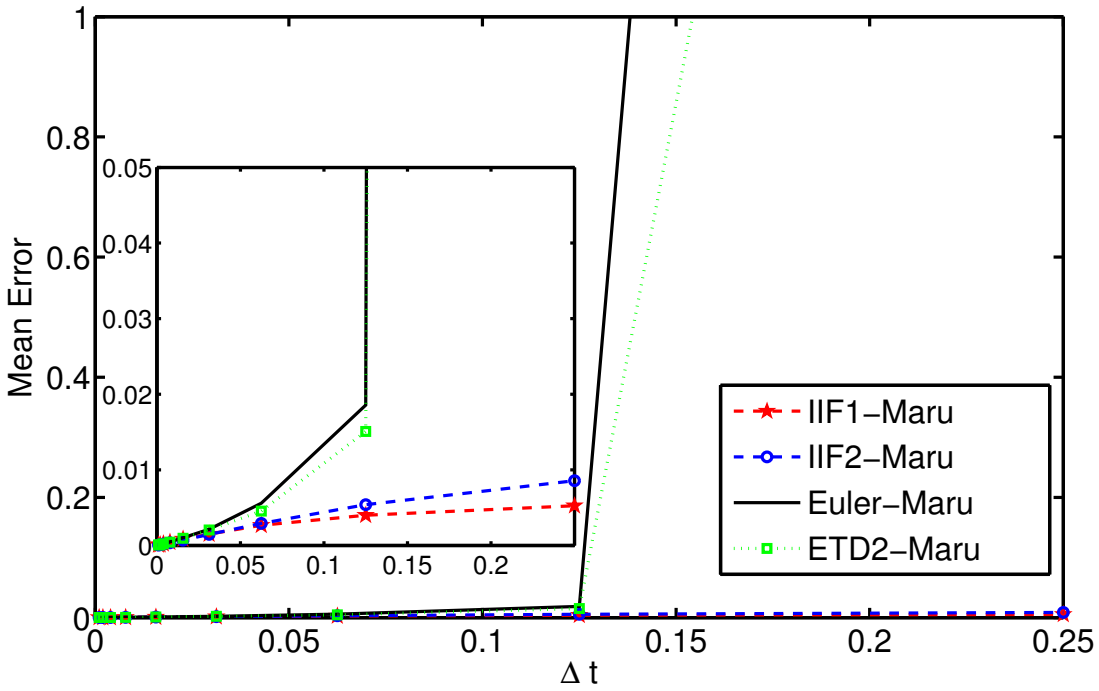

Figure 5.4: Comparison of the mean errors $\mathbb{E}|U_{\triangle t} - U_{\triangle t/2}|$ of the numerical solutions to Eq.(5.18) with additive noise obtained from the following methods: IIF1-Maruyama, IIF2-Maruyama, Euler Maruyama, and ETD2-Maruyama, when the reaction term is heavily stiff. The inserted figure shows the mean error comparison between different methods in more detail. The parameter values are as followed: $a = 1$, $b = -10$, and $\sigma = 0.1$.

Maruyama, and ETD2-Maruyama methods in different scenarios. Figure 5.5A and B represent the scenario where the noise amplitude is great compared to the magnitudes of the diffusion and reaction terms. Figure 5.5C and D represent the scenario where the magnitude of diffusion term is relatively large compared to those of the reaction and noise terms. In both scenarios, all the methods share an order of convergence of one as expected and no single method outperform the others. Figure 5.5 shows that both the IIF-Maruyama methods are comparable to other methods in terms of accuracy in the additive-noise case.

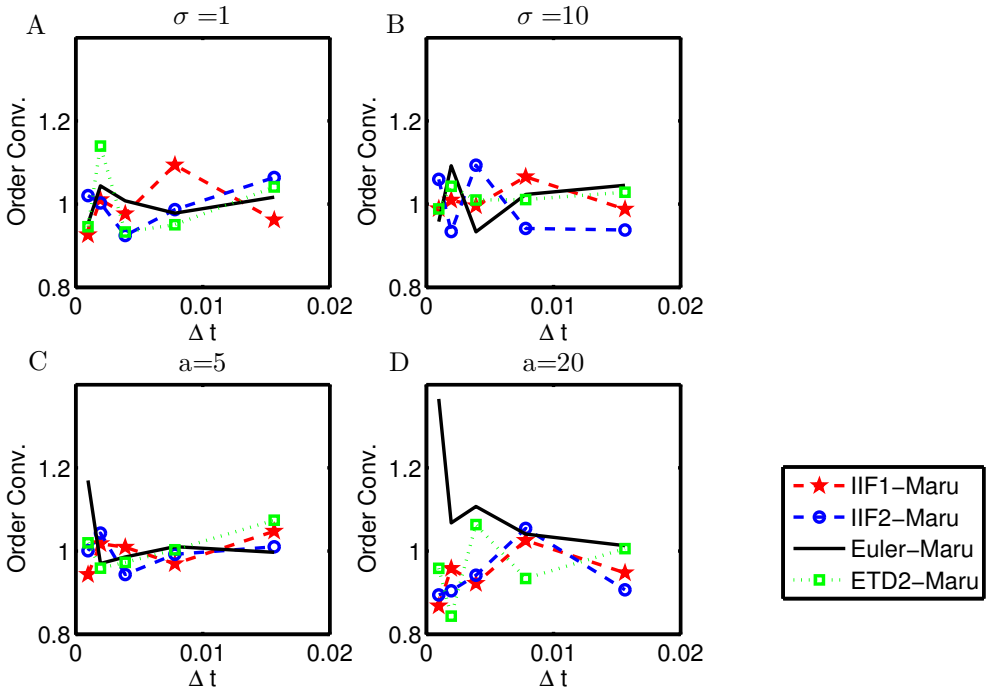

Figure 5.5: Comparison of the orders of convergence of the following methods: IIF1-Maruyama, IIF2-Maruyama, Euler Maruyama, and ETD2-Maruyama used to solve Eq.(5.18) with additive noise. Subplots A and B represent the orders of convergence of all the methods in the scenario that the noise amplitude is relatively large compared to the magnitudes of the diffusion and reaction terms, whose values are fixed to be $a = 0.1$ and $b = -1$. Subplots C and D represent the scenario where the magnitude of the diffusion term is great compared to those of the reaction and noise terms, whose values are fixed to be $\sigma = 0.1$ and $b = -1$.

Multiplicative Noise

For SODE with multiplicative noise, because the explicit solution for the linear stochastic ODE is known, the strong order of convergence can be estimated by the value γ if there exists a constant C such that:

$$\mathbb{E}|U_n - U(\tau)| \leq C\triangle t^{\gamma} \tag{5.41}$$

for any fixed $\tau = n\triangle t \in [0, T]$ where T is the final time and for $\triangle t$ sufficiently small. Let $T = L\triangle t$ for some time step $\triangle t$ [38]. At $\tau = T$, the order of convergence is calculated as the following

$$\text{Order of Convergence} = \frac{\log\left(\mathbb{E}|U_{L,\triangle t} - U(T)| \Big/ \mathbb{E}|U_{L,\triangle t/2} - U(T)|\right)}{\log 2} \tag{5.42}$$

where we denote $U_{L,\triangle t}$ to be the numerical solution at $t = T$ using time steps of size $\triangle t$ each and $U_{L,\triangle t/2}$ the numerical solution at time T using time steps of size $\triangle t/2$ each. To acquire the value for $\mathbb{E}|U_{L,\triangle t/2} - U(T)|$, we take the mean of $|U_{L,\triangle t/2} - U(T)|$ over 1000 independent Brownian paths, hence we call $\mathbb{E}|U_{L,\triangle t/2} - U(T)|$ the mean error of the numerical solution.

First, we test the accuracy of the IIF-Maruyama, Euler Maruyama, and ETD2-Maruyama methods when the magnitude of the reaction term is large and plot the mean errors of the numerical solutions for each method in Figure 5.6A. The mean error for the multiplicative case is defined to be $\mathbb{E}|U_n - U(\tau)|$ from Eq.(5.41). As in the additive case, when the reaction term is highly dominant, the IIF-Maruyama methods maintain low mean errors even when the time step $\triangle t$ is relatively large. From this figure, we observe that when $\triangle t$ is large enough, i.e $\triangle t = 1/8$, the mean errors of the solutions obtained from using Euler Maruyama and ETD2-Maruyama methods assume unreasonably large values and these two methods become unstable. At the same time step, both the IIF-Maruyama methods still maintain

stability, as evidenced by the reasonable mean errors resulted from the numerical solutions. We conclude that the IIF-Maruyama methods give reliable results in the case where the reaction-diffusion system is highly stiff in the reaction term.

Figure 5.6B shows that the orders of convergence for the IIF-Maruyama, Euler Maruyama, and ETD2-Maruyama methods are consistently one-half when the noise term is dominant. There is no real advantage of choosing one method over another in this scenario.

When the magnitude of the diffusion term is relatively large compared to those of the reaction and the noise terms, we notice that the mean errors $\mathbb{E}|U_n - U(\tau)|$ obtained from the IIF-Maruyama methods take much smaller values than those of the Euler Maruyama and ETD2-Maruyama methods, as evidenced by Figure 5.6C and D. Although all the methods have relatively small mean errors, the IIF-Maruyama methods have the smallest mean error values and therefore are more accurate than the other methods.

5.2.2 Tests on Stochastic Partial Differential Equations

Now we apply IIF-Maruyama methods to the following stochastic PDEs and compare the IIF methods with two other methods.

$$\frac{\partial U}{\partial t} = a\frac{\partial^2 U}{\partial x^2} + bU + g(U)\frac{\partial^2 W}{\partial x \partial t}. \tag{5.43}$$

where $0 \leq x \leq 1$ and $t \in [0, 0.125]$ along with a periodic boundary condition $U(0,t) = U(1,t)$.

We compare the orders of convergence from solving Eq.(5.43) in different scenarios among the following schemes: First-Order IIF-Maruyama method (5.16), Second-Order IIF-Maruyama method (5.17), Implicit-Euler Maruyama method, and Crank-Nicolson Maruyama method.

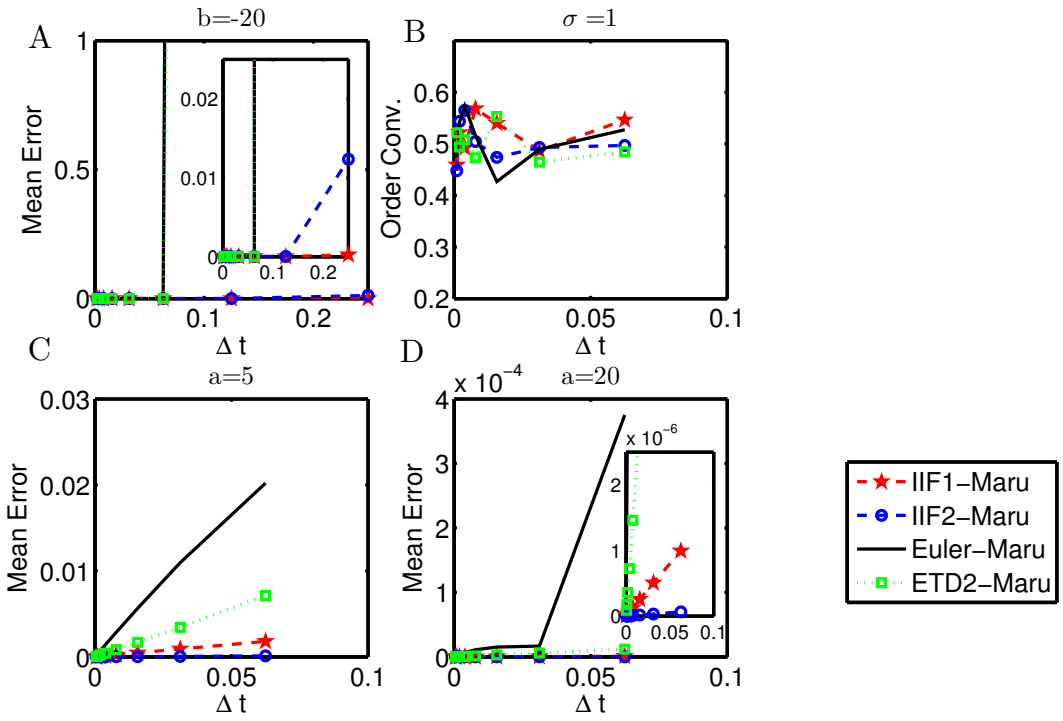

Figure 5.6: Comparison of the mean errors and orders of convergence between the following methods: IIF1-Maruyama, IIF2-Maruyama, Euler Maruyama, and ETD2-Maruyama, used to solve Eq.(5.18) with multiplicative noise. Subplot A is the comparison of the mean errors $\mathbb{E}|U_n - U(\tau)|$ of the numerical solutions when the reaction term is heavily stiff. The parameters used here are $a = 1$, $b = -20$, and $\sigma = 0.1$. In the scenario that the noise amplitude is relatively large compared to the magnitudes of the diffusion and reaction term, all methods display similar orders of convergence, as seen in subplot B. The parameter values for this subplot are $a = 0.1$, $b = -0.02$, and $\sigma = 1$. Finally, subplots C and D compare the mean errors $\mathbb{E}|U_n - U(\tau)|$ of the numerical solutions when the diffusion term is dominant using fixed parameters $\sigma = 0.1$ and $b = -2$. The time span for the simulations in subplot A is one and the time span used in subplots B-D is one-half. For subplots A and D, the inserted images show the mean errors of each method in more detail.

The Implicit-Euler Maruyama method is constructed in the same way as the Euler Maruyama method with the deterministic term being approximated using the backward-Euler method of order one. The overall order of convergence for this method is 1/2 due to the explicit approximation of the stochastic term as in the case of the Euler Maruyama method. Letting M denote the diffusion matrix after applying finite difference to Eq.(5.43), the method takes the following form,

Implicit-Euler Maruyama method

$$U_{j+1} = U_j + aM\triangle tU_{j+1} + \triangle t f(U_{j+1}) + g(U_j)\frac{\sqrt{\triangle t}\bar{\xi_j}}{\sqrt{\triangle x}}. \tag{5.44}$$

The construction of Crank-Nicolson method is also similar to that of the Euler Maruyama method, with the exception of the approximation of the deterministic term using the Crank-Nicolson method. For the same reason as the implicit-Euler Maruyama method, the overall order of convergence remains 1/2 for this method:

Crank-Nicolson Maruyama method

$$U_{j+1} = U_j + \frac{aM\triangle t}{2}(U_j + U_{j+1}) + \frac{1}{2}\triangle t(f(U_j) + f(U_{j+1})) + g(U_j)\frac{\sqrt{\triangle t}\bar{\xi_j}}{\sqrt{\triangle x}}. \tag{5.45}$$

We do not show the numerical results of the explicit Euler Maruyama scheme Eq.(5.33) and ETD2-Maruyama scheme Eq.(5.35) due to their disadvantages in stability and the associated computational cost.

To compute the order of convergence for each scheme mentioned, we use five different values of the number of spatial steps: N_1, N_2,...N_5 where N_1 is a power of 2 and $N_{i+1} = 2N_i$ for $i = 1, ..., 4$, and let the time step $\triangle t = 1/(4N_i)$. The solutions are numerically calculated over the time frame $[t_0, t_f]$ and generated over m different realizations of the Brownian sheet

$W(t,x)$. More information on how the Brownian sheet is generated can be found in [25]. We record the difference at the final time between two solutions obtained using N_i and N_{i+1} spatial steps and store this difference under the variable S_i for $i = 1, ...4$. The difference is the sum over m realizations of the sum of squared differences of the approximated solutions over N_1 spatial points, which are common to all solutions. Therefore, we obtain

$$S_i = \sum_{j=1}^{m} \sum_{k=1}^{N_1} \left(U_{j,k}^{i} - U_{j,k}^{i+1} \right)^2, \tag{5.46}$$

which offers a mean to calculate the numerical error of the scheme. Note that $U_{j,k}^{i}$ indicates the approximated solution at space step $x_k = k/N_1$ and at final time, which is obtained from using N_i spatial steps and j^{th} independent realization of the Brownian sheet [18]. Then the order of convergence can be estimated by $\log_2 R$ where the ratio $R = S_i/S_{i+1}$. When using this method of computing the order of convergence, both Implicit-Euler Maruyama and Crank-Nicolson Maruyama schemes converge with an order of $1/2$ for both additive and multiplicative noises [18]. As a result, we will use $1/2$ as the standard value of the order of convergence in the subsequent numerical comparisons. In this sub-section, we fix $m = 100$. From our experimentation, increasing the value of m has no effects on the orders of the convergence of each scheme. However, the values of $\{S_i\}_{i=1}^{4}$ will increase since these quantities depend on the value of m. In our tests, when $m = 500$, the values of $\{S_i\}_{i=1}^{4}$ are roughly five times larger than those obtained with $m = 100$. Similarly, if we increase m to 1000, $\{S_i\}_{i=1}^{4}$ are about ten times larger than their corresponding values when $m = 100$.

To compare the orders of convergence, we observe the following scenarios with both noises: how the order of convergence for each scheme is affected when the degree of stiffness increases, and whether each method still performs satisfactorily with large noise amplitude.

Stiff reaction

Figure 5.7A and B plots the values $\{S_i\}_{i=1}^4$ obtained from the numerical solutions of Eq.(5.43) with both additive and multiplicative noises versus the size of the space step when the reaction term is stiff with respect to the diffusion and noise terms. The numerical methods that are applied here are the IIF1-Maruyama, IIF2-Maruyama, Implicit-Euler Maruyama, and Crank-Nicolson Maruyama methods. For both types of noise, all methods converge with a rate of approximately 1/2. The Implicit-Euler Maruyama scheme along with the two IIF-Maruyama schemes have an advantage over the Crank-Nicolson Maruyama scheme when the values $\{S_i\}_{i=1}^4$ are considered.

In the interest of demonstrating the effectiveness of the IIF-Maruyama methods, we compare the orders of convergence between all of the aforementioned methods Eqs.(5.16), (5.17), (5.44), and (5.45) when the reaction term is extremely stiff for both additive and multiplicative noises. When noise is additive, all methods have an order of convergence of 1/2. However, the Implicit-Euler Maruyama and the two IIF-Maruyama methods perform much better than the Crank-Nicolson Maruyama method in terms of the mean errors $\{S_i\}_{i=1}^4$, as seen in Figure 5.7C. In Figure 5.7D, for multiplicative noise, the Implicit-Euler Maruyama and the IIF1-Maruyama methods converge at a much faster rate than 1/2 while the Crank-Nicolson Maruyama and the IIF2-Maruyama schemes maintain the 1/2 order of convergence. When using fewer number of space steps, the values S_i obtained from the Implicit-Euler Maruyama and IIF1-Maruyama methods are not as good as those of the other methods due to the large convergence rate. For example, when the spatial step sizes are 1/64, 1/128, and 1/256, the values S_1 and S_2 obtained from the IIF1-Maruyama and Implicit-Euler Maruyama methods are larger than those of the IIF2-Maruyama and Crank-Nicolson Maruyama methods. On the other hand, the IIF2-Maruyama method consistently has the smallest values for $\{S_i\}_{i=1}^4$, making it the most desirable method for solving a stochastic partial differential equation with an extremely stiff reaction term whose noise term can be either additive or multiplicative.

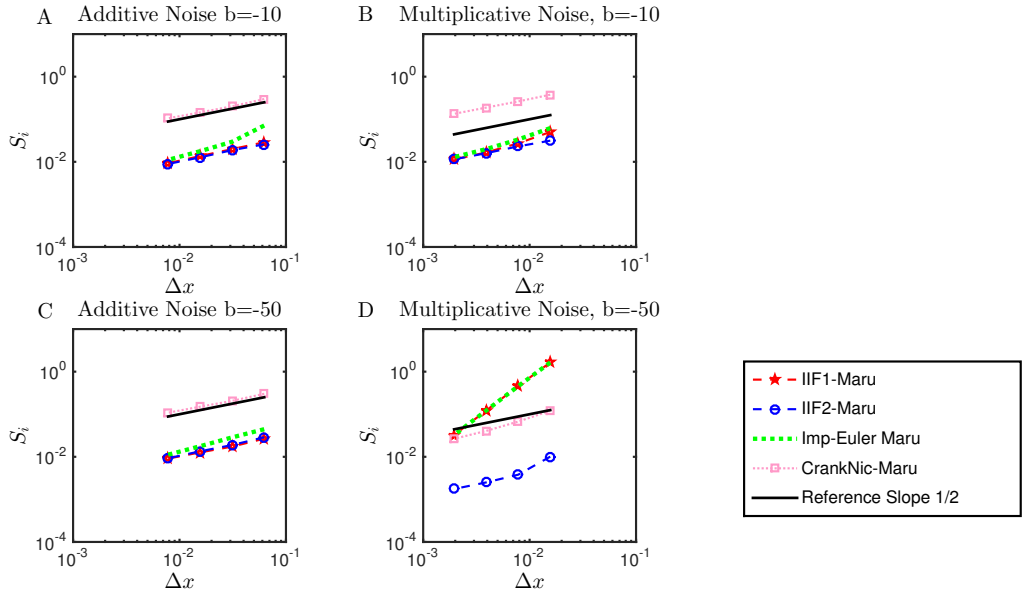

Figure 5.7: Comparison of the values $\{S_i\}_{i=1}^4$ from Eq.(5.46) and orders of convergence of the following methods: IIF1-Maruyama, IIF2-Maruyama, Implicit-Euler Maruyama, and Crank-Nicolson Maruyama used to solve Eq.(5.43) with both additive and multiplicative noises in the scenario where the reaction term is heavily stiff. Subplots A and B show the plots of $\{S_i\}_{i=1}^4$ of all the methods when the reaction term $b = -10$. Subplots C and D display plots of $\{S_i\}_{i=1}^4$ of all the methods when $b = -50$. For all the subplots, the values of the reaction and noise terms are fixed to be $a = 1$ and $\sigma = 0.1$. In subplots A and B, the time frame is chosen to be $t \in [0, 0.125]$ and for the remaining two subplots, $t \in [0, 0.025]$. Also, each plot contains the reference line of slope one-half for the purpose of order of convergence comparison.

Strong diffusion

We obtain the numerical solutions for Eq.(5.43) using the four methods when the diffusion coefficient is large. Then we compare the values $\{S_i\}_{i=1}^4$ from Eq.(5.46) and the orders of convergence of all methods by plotting $\{S_i\}_{i=1}^4$ versus the size of the space step along with a reference line of slope one-half in Figure 5.8A and B. The Crank-Nicolson Maruyama method is slightly more stable than the other methods for some large values of spatial step size but does not maintain this stability if the space step assumes a larger value than those shown in this figure. Both the IIF-Maruyama methods and the Implicit-Euler Maruyama method achieve the best values for $\{S_i\}_{i=1}^4$, while the Crank-Nicolson Maruyama method has significantly larger $\{S_i\}_{i=1}^4$ compared to them. When the spatial step size assumes a small-enough value, the IIF-Maruyama schemes and the Implicit-Euler Maruyama schemes have comparable orders of convergence with that of the Crank-Nicolson Maruyama scheme. With both the mean errors and orders of convergence taken into consideration, it is more advantageous to choose the IIF-Maruyama methods and the Implicit-Euler Maruyama method over the Crank-Nicolson Maruyama scheme.

Large noise amplitude

Figure 5.8C and D contains similar plots to Figure 5.8A and B in the case where Eq.(5.43) has a large noise term. For both additive and multiplicative noises, all methods have an order of convergence of one-half. Also, for all methods, the values of $\{S_i\}_{i=1}^4$ are slightly larger than the corresponding values obtained when the reaction term or diffusion term is stiff. Since the calculation of $\{S_i\}_{i=1}^4$ contains a double sum, the magnitude of $\{S_i\}_{i=1}^4$ could become quite large. Taking this into consideration, when the noise term has large magnitude, the IIF-Maruyama methods and the Implicit-Euler Maruyama method outperform the Crank-Nicolson Maruyama method significantly in terms of mean errors and thus are preferred.

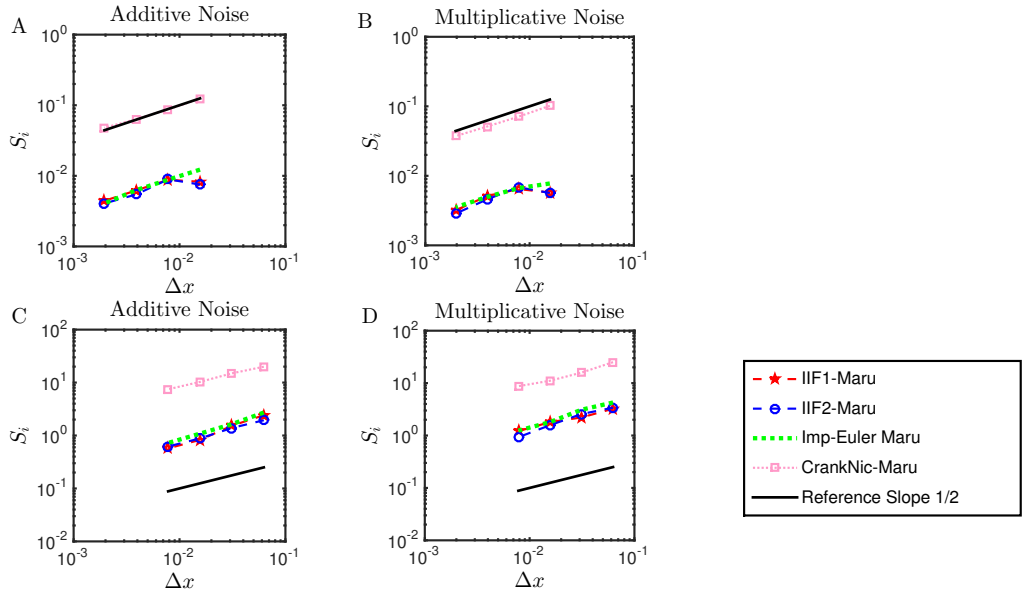

Figure 5.8: Comparison of the values $\{S_i\}_{i=1}^4$ from Eq.(5.46) and orders of convergence of the following methods: IIF1-Maruyama, IIF2-Maruyama, Implicit-Euler Maruyama, and Crank-Nicolson Maruyama used to solve Eq.(5.43). Subplots A and B are the plots of $\{S_i\}_{i=1}^4$ of all the methods when the diffusion term is stiff. The values of the diffusion, reaction, and noise terms are fixed to be $a = 20$, $b = -1$, and $\sigma = 0.1$ for all subplots. Subplots C and D display plots of $\{S_i\}_{i=1}^4$ of all the methods when the noise term assumes a large value. For these subplots, the values of the diffusion, reaction, and noise terms are fixed to be $a = 2$, $b = -1$, and $\sigma = 1$. In this figure, all the simulations are run for 0.125 time units. Also, each plot contains the reference line of slope one-half for the purpose of order of convergence comparison.

5.3 A Turing patterning system with noise

5.3.1 1-dimensional activator-substrate system

Finally we apply the IIF2-Maruyama scheme to a Turing patterning system that contains noise to study how noise may affect the formation of patterns. We use the activator-substrate system as an example, whose non-dimensional form is as followed [56, 71] :

$$\frac{\partial A}{\partial t} = D\frac{\partial^2 A}{\partial x^2} + SA^2 - A + \rho \tag{5.47}$$

$$\frac{\partial S}{\partial t} = \frac{\partial^2 S}{\partial x^2} + \mu(1 - SA^2). \tag{5.48}$$

The boundary conditions are no-flux. The constant D measures the diffusion coefficient ratio of activator to substrate. The parameters ρ and μ measures the production rates of activator and substrate, respectively. This system has known homogeneous solutions:

$$A^* = 1 + \rho, \quad S^* = (1 + \rho)^{-2} \tag{5.49}$$

One of the features of the activator-susbtrate system Eqs.(5.47) and (5.48) is that the substrate S can be consumed during the autocatalysis of activator A and the interactions between the activator and substrate of this system lead to the formation of spatially inhomogeneous patterns [56, 71]. Without noise, the inhomogeneous steady state patterns may be obtained in the parameter ranges

$$\rho \in (0, 1), \quad \mu \in \left(\frac{2}{1+\rho} - 1, 2\right), \quad D \in \left(0.001, \frac{1}{\mu}\left(\sqrt{\frac{2}{1+\rho}} - 1\right)\right). \tag{5.50}$$

The initial guesses for the pattern (A_0, S_0) are chosen as [71]

$$A_0 = A^*(1 + \gamma\delta_A(x)), \quad S_0 = S^*(1 + \gamma\delta_S(x)) \tag{5.51}$$

where $\delta_A(x)$ and $\delta_S(x)$ are standard normally-distributed random variables with zero mean and variance one. Different patterns can arise from a slight variations in initial conditions [77]. If we choose the following parameters,

$$t \in [0, 101],\ x \in (0, 10),\ \rho = 0.01,\ \mu = 1,\ D = 0.1,\ \gamma = 0.02,\ dx = 10/2^7,\ dt = dx/4. \tag{5.52}$$

the different permutations of our initial conditions in Eq.(5.51) result in six distinct combinations of inhomogeneous steady state patterns for solutions A and S, similar to those in Figure 5.9.

Now, we add multiplicative noise to the Eqs.(5.47) (5.48) to obtain:

$$\frac{\partial A}{\partial t} = D\frac{\partial^2 A}{\partial x^2} + SA^2 - A + \rho + \epsilon_A A\frac{\partial^2 W}{\partial x \partial t}, \tag{5.53}$$

$$\frac{\partial S}{\partial t} = \frac{\partial^2 S}{\partial x^2} + \mu(1 - SA^2) + \epsilon_S S\frac{\partial^2 W}{\partial x \partial t}. \tag{5.54}$$

Results

We implement the IIF2-Maruyama method described in Eq.(5.15) to Eqs.(5.53) and (5.54) and examine the changes to the deterministic steady-state patterns when a multiplicative noise is added to both Eqs.(5.47) and (5.48). For our implementation, the parameters from

Eq.(5.52) are selected.

We want the time span for our simulations to be sufficiently long so that we can observe the long term behaviors of the solutions A and S. More specifically, in our cases, we want to see whether the long-term solutions assume any patterns observed in their deterministic steady-state counterparts. Unlike the previous study for the deterministic equations [71] in which fluctuations in the initial conditions are critically important in generating the patterns, we fix the initial conditions $A_0 = A^*$ and $S_0 = S^*$ defined in Eq.(5.49).

Interestingly, we obtain similar patterns in spite of the uniform initial conditions when the relatively small values to the noise coefficients, i.e $\epsilon_A = 0.005$ and $\epsilon_S = 0.01$, are given. The six different combinations of patterns for the long time solutions A and S that exist can be seen in Figure 5.9. We note that these six different combinations of patterns are the same inhomogeneous steady state patterns obtained from solving the deterministic equations Eqs.(5.47) and (5.48) [71]. Hence, adding multiplicative noise to the activator-substrate system enables us to obtain the inhomogeneous steady-state patterns that are otherwise obtained through the fluctuations of the initial conditions, as previously predicted [77].

Because each stochastic solution may reach a different steady state even with the same initial conditions in a deterministic form, we perform 100, 500, and 1000 simulations to see which combination of patterns shows up more frequently. In Table 5.1, we fix the initial conditions as in Eq.(5.49) and choose the spatial step size to be $dx = 10/2^6$ and $dx = 10/2^7$. This change in dx does not affect the frequency of presence of each combination of patterns. For Table 5.2, the values of the initial conditions are randomly permuted as in Eq.(5.51) for each independent path, while dx is fixed at $10/2^7$. From these two tables, we see that the first combination of patterns is consistently the most favored type of patterns, with a $> 30\%$ chance of occurrence, with the second combination of patterns being the second most typical combination of patterns. In addition, the frequency of appearance of each type of patterns is independent of the effects of extra fluctuations on the initial conditions, as evidenced by Table

	Percentage of appearance					
Spatial step size	$dx = 10/2^6$			$dx = 10/2^7$		
No. simulations	100	500	1000	100	500	1000
Pattern 1	35%	33.4%	34.4%	39%	35.8%	36.7%
Pattern 2	28%	26.4%	29%	25%	22.8%	26.1%
Patterns 3 and 5	23%	19.8%	18.9%	15%	20.2%	18.8%
Patterns 4 and 6	14%	20.4%	17.7%	21%	21.2%	18.4%

Table 5.1: Percentage of occurrence of each combination patterns for 100, 500, and 1000 different simulations. The initial conditions are uniformly fixed as in Eq.(5.49).

	Percentage of appearance		
No. simulations	100	500	1000
Pattern 1	41%	40.2%	37%
Pattern 2	30%	26.4%	26.7%
Patterns 3 and 5	15%	15.2%	18.1%
Patterns 4 and 6	14%	18.2%	18.2%

Table 5.2: Percentage of occurrence of each combination patterns for 100, 500, and 1000 different simulations. In this table, $dx = 2^{-7}$ and the initial conditions are permuted as in Eq.(5.51).

5.2. This implies that the first combination of patterns is likely to make up the standard type of patterns that the long-term activator and substrate solutions are supposed to assume. The lack of robustness in pattern formation of the activator and substrate levels can be improved by adding growth factor to the system, in particular apical growth in the case of intrinsic noise [77]. As the domain grows, the space between the activated regions (characterized by the activator maxima or the substrate minima) is enlarged while the substrate concentration is quickly diffusing and increasing. This increase in substrate prompts a higher production of the activator at the side of the maxima in comparison to its center, resulting in the movement of the activator maxima to regions with higher substrate concentration [56]. Attributing the appropriate type of growth to the system can help stabilize the pattern formation over time where the robust patterns for both activator and substrate levels are the first combination of patterns.

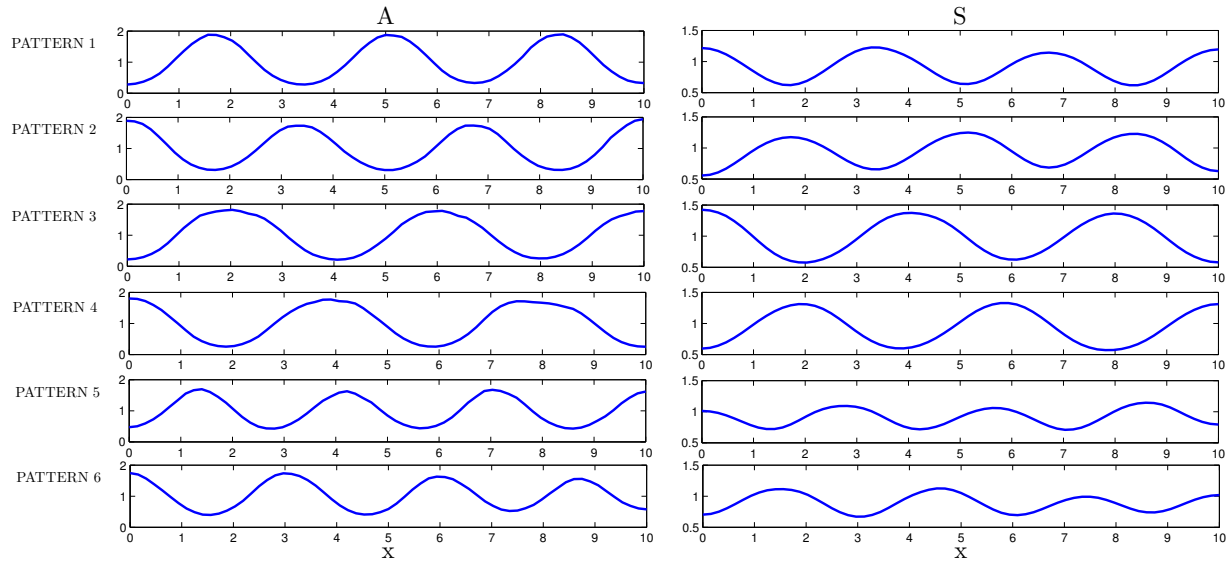

Figure 5.9: The six different combinations of steady-state patterns for long-term solutions A and S to the one-dimensional stochastic system in Eqs.(5.53) and (5.54). The values for the noise coefficients are $\epsilon_S = 0.01$ and $\epsilon_A = 0.005$.

Computational efficiency

Next, we discuss the computational efficiency of the two IIF-Maruyama methods used to solve the Activator-Substrate system with multiplicative noise, Eqs.(5.53) and (5.54), by comparing their performances with that of an explicit method, which we choose to be the Euler Maruyama method. We keep all the parameter values as described previously Eq.(5.52) with the exception of time, which is changed to $t \in [0, 1]$. In Table 5.3, the mean errors, the order of convergence, and the computational time for each method are recorded. To calculate the mean errors and the order of convergence, we carry out 100 different simulations and apply Eq.(5.46). We choose $dt = dx/2$ for the IIF1-Maruyama method, $dt = dx$ for the IIF2-Maruyama method, and $dt = (dx)^2/3$ for the Euler Maruyama method to ensure convergence. In this table, we denote N to be the number of spatial grid points that partition the interval (0,10). Besides $N = 2^3$, all methods display similar orders of convergence and mean errors of similar magnitude. Next, we discuss the computational time in seconds of each method. Each time listed in our table is the total time each method takes to compute the solution over 100

different Brownian paths. For the computational time of both IIF-Maruyama methods, we include the time it takes to calculate the exponential matrix using the Matlab function *expm*. We observe that with coarser spatial grids, i.e when $N = 2^3$ and $N = 2^4$, the Euler Maruyama method surpasses both IIF-Maruyama methods in terms of computational effort. With finer grids, i.e $N = 2^6$ and 2^7, it takes the Euler Maruyama method twice as long as it takes the IIF2-Maruyama method in computing the solutions over 100 Brownian paths. Similarly, the IIF1-Maruyama method is put at a disadvantage with respect to computational speed when the grids are coarse but quickly catches up to the Euler Maruyama method with refined grids. We note that the second-order IIF-Maruyama method is more efficient than the first-order IIF-Maruyama method. In addition, the second-order IIF-Maruyama method catches up to the Euler Maruyama method much faster in improving its computational speed, which is demonstrated by the similar speeds between these two methods when $N = 2^4$. Meanwhile, the computational speed of the IIF1-Maruyama method does not catch up to that of the Euler Maruyama method until $N = 2^5$. In short, due to the restriction of the temporal step size that is required to maintain numerical stability, the Euler Maruyama scheme is less computationally efficient than the IIF1-Maruyama and the IIF2-Maruyama schemes when a finer spatial grid is required. Between the two IIF-Maruyama methods, the second-order IIF-Maruyama method is more desirable for its computational efficiency than the first-order IIF-Maruyama method.

5.3.2 2-dimensional activator-substrate system

For two or three-dimensional systems, direct application of IIF is costly because the storage and computation of the exponential matrix $e^{aM\triangle t}$ in the IIF methods may become very large. Similarly to solving the deterministic systems in two or three dimensions, here we use the compact integration factor methods (cIIF) [82], in which the discretized diffusion operator is represented in a compact form that requires storage only proportional to the

	IIF1-Maruyama			IIF2-Maruyama			Euler Maruyama		
	$dt = dx/2$			$dt = dx$			$dt = (dx)^2/3$		
N	S	Order	Time(s)	S	Order	Time(s)	S	Order	Time(s)
2^3	3.52e-2	2.36	0.158	1.27e-2	0.61	0.014	1.41e-2	0.28	0.009
2^4	6.88e-3	0.65	0.154	8.35e-3	0.62	0.101	1.02e-2	0.47	0.051
2^5	4.38e-3	0.37	0.233	5.42e-3	0.69	0.194	6.74e-3	0.60	0.230
2^6	3.38e-3	0.93	0.614	3.35e-3	0.86	0.481	3.70e-3	0.86	1.093
2^7	1.77e-3	0.58	2.019	1.84e-3	0.68	1.208	2.14e-3	0.79	4.431

Table 5.3: The mean errors S defined in Eq.(5.46), orders of convergence, and computational cost when solving the activator-substrate system with noise described in Eqs.(5.53) and (5.54) obtained by each method: IIF1-Maruyama, IIF2-Maruyama, and Euler Maruyama. The results are computed over 100 Brownian paths with N denoting the number of spatial grid points that partition the interval (0,10).

number of unknowns instead of the square of the number of unknowns in the case of non-compact IIF methods for the exponentials of matrices. cIIF methods can be combined with the Maruyama method in the same manner as the integration of the IIF methods with the Maruyama method, and both the cIIF and the IIF methods share the same desirable stability properties [82].

Here we construct the cIIF-Maruyama methods by estimating the deterministic diffusion and reaction terms using the cIIF methods and the stochastic term using the explicit Maruyama approximation. To demonstrate the efficiency of the cIIF-Maruyama methods, we apply the cIIF2-Maruyama method to the two-dimensional version of the activator-substrate system with no-flux boundary conditions Eqs.(5.53) and (5.54) presented in the previous section. Similarly to the one-dimensional case, we compute the solutions over the space $(0, 10)\times(0, 10)$ using the steady states in Eq.(5.49) as the initial conditions. The time window for the simulation is set to be $t \in [0, 200]$ and $dt = dx$ where $dx = dy = 2^{-7} \times 10$. The rest of the parameter values in Eq.(5.52) remain the same. Figure 5.10 displays the contour plot of one of the final patterns of the solutions A and S .

The computational time it takes for the cIIF2-Maruyama method to compute one stochastic

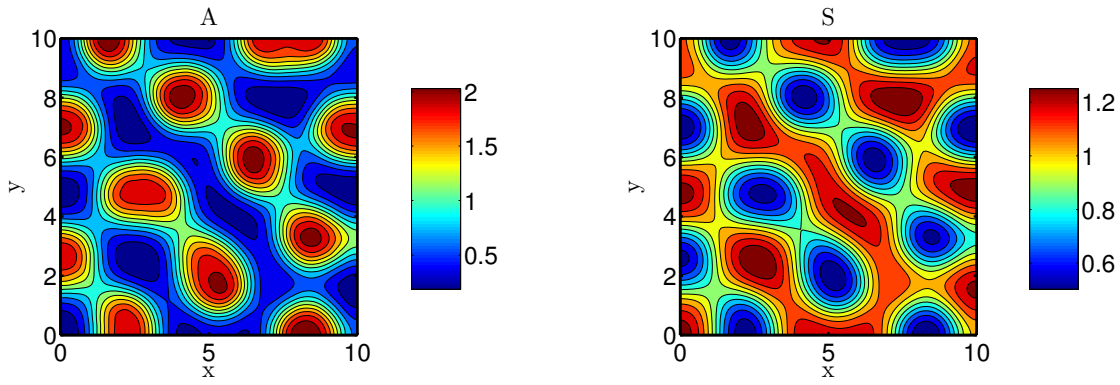

Figure 5.10: One of the final patterns obtained for long-term solutions A and S when solving the two-dimensional stochastic system in Eqs.(5.53) and (5.54). The values for the noise coefficients are $\epsilon_S = 0.01$ and $\epsilon_A = 0.005$.

solution to the two-dimensional activator-substrate system is 126.789 seconds for this case. If we use the Euler-Maruyama method to solve this system using the same spatial step size, it takes 206.740 seconds due to the restriction on the temporal step size, which we set to be $dt = (dx)^2/5$. In summary, due to the stability property, the cIIF-Maruyama methods are as efficient in solving two dimensional stochastic reaction-diffusion systems as the one-dimensional systems.

5.4 Conclusions and Discussion

By convention, stochastic stiffness is defined to be the result of the different time scales caused by the large discrepancies in the magnitudes of the Lyapunov exponents [8, 54]. For that reason, stochastic stiffness can occur in the deterministic term, stochastic term, or both. When solving a stochastic differential equation, the problem of stiffness that stems from the stochastic term has been studied previously [1, 8, 9, 37, 54, 65, 80, 107]. Here, we have

focused on the treatment of stiffness of the reaction term for a stochastic reaction-diffusion system. By taking advantage of the existing semi-implicit integrating factor method that is both computationally efficient and absolutely stable at solving stiff deterministic reaction-diffusion systems, we have developed a new class of temporal schemes for reaction-diffusion systems with both additive and multiplicative noises. Similarly to the deterministic case, the new numerical schemes presented remove the restriction imposed on the temporal step size by the linear diffusion term by treating this term exactly while dealing with the stiff reaction term through an implicit approximation. Numerical comparisons show that the construction using the IIF technique to approximate the deterministic term allows the new methods to achieve better stability and good efficiency. While the explicit treatment of the diffusion in IIF naturally leads to good approximations on strong diffusion, the new IIF-Maruyama methods mainly offer an efficient approach to deal with stiff reactions in a reaction-diffusion systems. In general, this method is mostly effective when the reactions are very stiff while diffusion is still important in a stochastic reaction-diffusion system.

The approach used here in combining IIF for reaction and diffusion and Maruyama for the stochastic terms can be adapted in a straightforward fashion for compact IIF (cIIF) methods [82], which is effective for systems in two or three dimensions. With the compact representation for the differential operators, it would be more efficient in simulating 2D and 3D systems using cIIF than IIF-Maruyama. Another improvement on IIF-Maruyama is its order of convergence.

In the case of multiplicative noise, the order of convergence of the IIF-Maruyama methods is confined to a value of one-half. With the Euler Maruyama method, the order of convergence can be improved by adding more terms from the Ito-Taylor expansion to construct methods with higher order of accuracy such as the Milstein method [54]. Integrating such approach with IIF method might lead to higher order of accuracy with similar stability property of IIF-Maruyama. However, because the diffusion term is not well-defined for the Brownian sheet

due to a lack of a well-defined spatial derivative for the Wiener process, a direct application of the Milstein method to stochastic PDEs may only lead to half-order of convergence. To deal with this difficulty, one might need to use the Q-Wiener process instead of the Brownian sheet to approximate the diffusion [74]. Similarly to the IIF methods, which can be used for the deterministic systems in various forms, spatial dimensions, and in combination with other approaches for treating additional terms such as convections, IIF-Maruyama methods may have broad applications in simulating stochastic partial differential equations in various forms and containing stiff reactions.

Bibliography

[1] Assyr Abdulle. Explicit methods for stiff stochastic differential equations. In B. Engquist, O. Runborg, and Y. Tsai, editors, *Numerical Analysis of Multiscale Computations*, volume 82 of *Lecture Notes in Computational Science and Engineering*, pages 1–22. Springer Berlin Heidelberg, 2012.

[2] Y. H. Ahn and D. L. Gibbons. ZEB1 drives prometastatic actin cytoskeletal remodeling by downregulating miR-34a expression. *The Journal of Clinical Investigation*, 122, 2012.

[3] J. Baulida and A. Garcia de Herreros. Snail1-driven plasticity of epithelial and mesenchymal cells sustains cancer malignancy. *Biochim Biophys Acta*, 1856(1):55–61, 2015.

[4] A. Biddle, X. Liang, L. Gammon, B. Fazil, L. Harper, and et al. Cancer stem cells in squamous cell carcinoma switch between two distinct phenotypes that are preferentially migratory or proliferative. *Cancer Research*, 71(15):5317–5326, 2011.

[5] T. Brabletz, A. Jung, S. Reu, and et al. Variable beta-catenin expression in colorectal cancers indicates tumor progression driven by the tumor environment. *Proc Natl Acad Sci U S A*, 98:10356–10361, 2001.

[6] C. P. Bracken, P. A. Gregory, N. Kolesnikoff, A. G. Bert, and et al. A double-negative feedback loop between ZEB1-SIP1 and the microRNA-200 family regulates epithelial-mesenchymal transition. *Cancer Research*, 68:7846–7854, 2008.

[7] U. Burk, J. Schubert, U. Wellner, O. Schmalhofer, E. Vincan, and et al. A reciprocal repression between ZEB1 and members of the miR-200 family promotes EMT and invasion in cancer cells. *EMBO reports*, 9:582–589, 2008.

[8] K. Burrage, P. M. Burrage, and T. Tian. Numerical methods for strong solutions of stochastic differential equations: an overview. *Proc. R. Soc. Lond. A*, 460(2041):373–402, 2004.

[9] K. Burrage and T. Tian. The composite Euler method for stiff stochastic differential equations. *Jour. Comp. Applied Math.*, 131:407–426, 2001.

[10] A. Q. Cai, K. Radtke, A. Linville, and et al. Cellular retinoic acid-binding proteins are essential for hindbrain patterning and signal robustness in zebrafish. *Development*, 139:2150–2155, 2012.

[11] A. Cano, M. A. Pérez-Moreno, and I. Rodrigo. The transcription factor snail controls epithelial-mesenchymal transitions by repressing E-cadherin expression. *Nature Cell Biology*, 2:76–83, 2000.

[12] C. L. Chaffer, N. Marjanovic, T. Lee, G. Bell, C. G. Kleer, and et al. Poised chromatin at the Zeb1 promoter enables breast cancer cell plasticity and enhances tumorigenicity. *Cell*, 154(1):61–74, 2013.

[13] H. H. Chang, M. Hemberg, M. Barahona, and et al. Transcriptome-wide noise controls lineage choice in mammalian progenitor cells. *Nature*, 453:544–547, 2008.

[14] J. Chen, Q. Han, and D. Pei. EMT and MET as paradigms for cell fate switching. *J Mol Cell Biol*, 4:66–69, 2012.

[15] M. Chen, L. Wang, C. C. Liu, and et al. Noise attenuation in the ON and OFF states of biological switches. *ACS Synth Biol*, 2:587–593, 2013.

[16] C. M. Chuong and R. B. Widelitz. The river of stem cells. *Cell Stem Cell*, 4:100–102, 2009.

[17] R. Clewley. Hybrid models and biological model reduction with PyDSTool. *PLoS Comput. Biol.*, 8(e1002628), 2013.

[18] A. M. Davie and J. G. Gaines. Convergence of numerical schemes for the solution of parabolic stochastic partial differential equations. *Mathematics of Computation*, 70(233):121–134, 2000.

[19] T. S. Deisboeck, Z. Wang, P. Macklin, and V. Cristini. Multiscale cancer modeling. *Annu. Rev. Biomed. Eng.*, 13, 2011.

[20] A. Diaz-Lopez, J. Diaz-Martin, G. Moreno-Bueno, and et al. Zeb1 and Snail1 engage miR-200f transcriptional and epigenetic regulation during EMT. *International Journal of Cancer*, 136:E62–73, 2015.

[21] A. Eger, K. Aigner, S. Sonderegger, B. Dampier, and et al. DeltaEF1 is a transcriptional repressor of E-cadherin and regulates epithelial plasticity in breast cancer cells. *Oncogene*, 24:2375–2385, 2005.

[22] B. Ermentrout. *Simulating, Analyzing, and Animating Dynamical Systems: A Guide to XPPAUT for Researchers and Students*. Society for Industrial and Applied Mathematics, 2002.

[23] I. Th. Famelis, F. Xanthos, and G. Papageorgiou. Numerical solution of stochastic differential equations with additive noise by Runge-Kutta methods. *Journal of Numerical Analysis, Industrial and Applied Mathematics*, 4(3-4):171–180, 2009.

[24] P. Friedl and D. Gilmour. Collective cell migration in morphogenesis, regeneration, and cancer. *Nature Reviews, Molecular Cell Biology*, 10(7):445–457, 2009.

[25] J. G. Gaines. Numerical experiments with S(P)DE's. In A. M. Etheridge, editor, *Stochastic Partial Differential Equations*, volume 216 of *London Math. Soc. Lecture Note Ser.*, pages 55–71. Springer Berlin Heidelberg, 1995.

[26] S. Gaudet, S. L. Spencer, W. W. Chen, and et al. Exploring the contextual sensitivity of factors that determine cell-to-cell variability in receptor-mediated apoptosis. *PLoS Comput Biol*, 8, 2012.

[27] N. J. Godde, R. C. Galea, I. A. Elsum, and P. O. Humbert. Cell polarity in motion: redefining mammary tissue organization through EMT and cell polarity transitions. *Journal of Mammary Gland Biol/ Neoplasia*, 15:149–168, 2010.

[28] P. A. Gregory, C. P. Bracken, E. Smith, A. G. Bert, and et al. An autocrine TGFβ/ZEB/miR-200 signaling network regulates establishment and maintenance of epithelial-mesenchymal transition. *Molecular Biology of the Cell*, 22:1686–1698, 2011.

[29] A. Grosse-Wilde, A. F. D'Hérouël, E. McIntosh, G. Ertaylan, A. Skupin, and et al. Stemness of the hybrid epithelial/ mesenchymal state in breast cancer and its association with poor survival. *PLoS One*, 10(5):e0126522, 2015.

[30] S. Guaita, I. Puig, C. Franci, and et al. Snail induction of epithelial to mesenchymal transition in tumor cells is accompanied by MUC1 repression and ZEB1 expression. *The Journal of Biological Chemistry*, 277:39209–39216, 2002.

[31] D. Guo, B. Xu, X. Zhang, and M. M. Dong. Cancer stem-like side population cells in the human nasopharyngeal carcinoma cell line CNE-2 possess epithelial mesenchymal transition properties in association with metastasis. *Oncol. Rep.*, 28(1):241–247, 2012.

[32] P. B. Gupta, C. M. Fillmore, G. Jiang, and et al. Stochastic state transitions give rise to phenotypic equilibrium in populations of cancer cells. *Cell*, 146:633–644, 2011.

[33] I. Gyöngy and D. Nualart. Implicit scheme for stochastic parabolic partial differential equations driven by space-time white noise. *Potential Analysis*, 7:725–757, 1997.

[34] Y. Hart, Y. E. Antebi, A. E. Mayo, and et al. Design principles of cell circuits with paradoxical components. *Proc Natl Acad Sci U S A*, 109:8346–8351, 2012.

[35] E. Hausenblas. Approximation for semilinear stochastic evolution equations. *Potential Analysis*, 18:141–186, 2003.

[36] K. Hayashi, S. de Sousa Lopes, and et al F. Tang. Dynamic equilibrium and heterogeneity of mouse pluripotent stem cells with distinct functional and epigenetic states. *Cell Stem Cell*, 3:391–401, 2008.

[37] D. J. Higham. Mean-square and asymptotic stability of the stochastic theta method. *Siam J. Numer. Anal.*, 38(3):753–769, 2000.

[38] D. J. Higham. An algorithmic introduction to numerical simulation of stochastic differential equations. *SIAM Review*, 43(3):525–546, 2001.

[39] T. Hong, C. Oguz, and J. J. Tyson. A mathematical framework for understanding four-dimensional heterogeneous differentiation of CD4+ T cells. *Bulletin of Mathematical Biology*, pages 1–19, 2015.

[40] T. Hong, K. Watanabe, C. Ta, A. Villarreal-Ponce, Q. Nie, and X. Dai. An Ovol2-Zeb1 mutual inhibitory circuit governs bidirectional and multi-step transition between epithelial and mesenchymal states. *PLoS Comp. Bio*, 11(11):e1004569, 2015.

[41] T. Hong, J. Xing, L. Li, and J. J. Tyson. A simple theoretical framework for understanding heterogeneous differentiation of CD4+ T cells. *BMC Syst. Biol.*, 6(1):66, 2012.

[42] R. Y. Huang, M. K. Wong, T. Z. Tan, K. T. Kuay, A. H. Ng, and et al. An EMT spectrum defines an anoikis-resistant and spheroidogenic intermediate mesenchymal state that is sensitive to e-cadherin restoration by a SRC-kinase inhibitor, saracatinib (AZD0530). *Cell Death Dis.*, 4(e915), 2013.

[43] S. Huang. Genetic and non-genetic instability in tumor progression: link between the fitness landscape and the epigenetic landscape of cancer cells. *Cancer Metastasis Reviews*, 32:423–448, 2013.

[44] A. Jentzen and P. E. Kloeden. The numerical approximation of stochastic partial differential equations. *Milan J. Math.*, 77:205–244, 2009.

[45] A. Jentzen and P. E. Kloeden. Overcoming the order barrier in the numerical approximation of stochastic partial differential equations with additive space-time noise. *Proc. R. Soc. A*, 465(2102):649–667, 2009.

[46] D. Jia, M. K. Jolly, M. Boareto, and et al. OVOL guides the epithelial-hybridmesenchymal transition. *Oncotarget*, 6:1543615448, 2015.

[47] M. K. Jolly, M. Boareto, B. Huang, D. Jia, M. Lu, and et al. Implications of the hybrid epithelial-mesenchymal phenotype in metastasis. *Front. Oncol.*, 5(155), 2015.

[48] M. K. Jolly, B. Huang, M. Lu, S. A. Mani, and et al. Towards elucidating the connection between epithelial-mesenchymal transitions and stemness. *J. R. Soc, Interface*, 11(101), 2014.

[49] M. K. Jolly, D. Jia, M. Boareto, and et al. Coupling the modules of EMT and stemness: A tunable 'stemness window' model. *Oncotarget*, 6:25161–25174, 2015.

[50] R. Kalluri and R. A. Weinberg. The basics of epithelial-mesenchymal transition. *The Journal of Clinical Investigation*, 119:1420–1428, 2009.

[51] N. G. Van Kampen. *Stochastic Processes in Physics and Chemistry*. Elsevier, Amsterdam, 3 edition, 2007.

[52] J. Keizer. *Statistical Thermodynamics of Nonequilibrium Processes*. Springer-Verlag, 1987.

[53] N. H. Kim, H. S. Kim, X. Y. Li, and et al. A p53/miRNA-34 axis regulates Snail1-dependent cancer cell epithelial-mesenchymal transition. *The Journal of Cell Biology*, 195:417–433, 2011.

[54] P. E. Kloeden and E.Platen. *Numerical solutions of stochastic differential equations*. Springer, 1992.

[55] P. E. Kloeden and S. Shott. Linear-implicit strong schemes for Itô-Galerkin approximations of stochastic PDEs. *Jour. Applied Math. Sto. Anal.*, 14:47–53, 2001.

[56] A. J. Koch and H. Meinhardt. Biological pattern formation: from basis mechanisms to complex structures. *Rev. Mod. Phys.*, 66:1481–1507, 1994.

[57] S. Kondo and T. Miura. Reaction-diffusion model as a framework for understanding biological pattern formation. *Science*, 329(5999):1616–1620, 2010.

[58] R. Kubo. The fluctuation-dissipation theorem. *Reports on Progress in Physics*, 29:255–284, 1966.

[59] D. A. Lawson, N. R. Bhakta, K. Kessenbrock, and et al. Single-cell analysis reveals a stem-cell program in human metastatic breast cancer cells. *Nature*, 526:131–135, 2015.

[60] M. J. Lawson, B. Drawerr, M. Khammash, L. Petzold, and T. Yi. Spatial stochastic dynamics enable robust cell polarization. *PLoS Comput Biol*, 9(7), 2013.

[61] B. Lee, A. Villarreal-Ponce, M. Fallahi, J. Ovadia, and et al. Transcriptional mechanisms link epithelial plasticity to adhesion and differentiaton of epidermal progenitor cells. *Dev Cell*, 29:47–58, 2014.

[62] J. Lei, S. A. Levin, and Q. Nie. Mathematical model of adult stem cell regeneration with cross-talk between genetic and epigenetic regulation. *Proc Natl Acad Sci U S A*, 111:880–887, 2014.

[63] C. Li and J. Wang. Quantifying cell fate decisions for differentiation and reprogramming of a human stem cell network: Landscape and biological paths. *PLOS Computational Biology*, 9(8):1–14, 2013.

[64] C. Li and J. Wang. Quantifying the landscape for development and cancer from a core cancer stem cell circuit. *Cancer Research*, 75(13):2607–2618, 2015.

[65] T. Li, A. Abdulle, and E. Weinan. Effectiveness of implicit methods for stiff stochastic differential equations. *Commun. Comput. Phys.*, 3(2):295–307, 2008.

[66] W. A. Lim, C. M. Lee, and C. Tang. Design principles of regulatory networks: searching for the molecular algorithms of the cell. *Mol Cell*, 49:202–212, 2013.

[67] S. Liu, Y. Cong, D. Wang, Y. Sun, L. Deng, and et al. Breast cancer stem cells transition between epithelial and mesenchymal states reflective of their normal counterparts. *Stem Cell Reports*, 2(1):78–91, 2014.

[68] X. Liu, S. Johnson, and et al S. Liu. Nonlinear growth kinetics of breast cancer stem cells: implications for cancer stem cell targeted therapy. *Sci Rep*, 3, 2013.

[69] X. Liu and Q. Nie. Compact integration factor methods for complex domains and adaptive mesh refinement. *Jour. Comp. Physics*, 229:5692–5706, 2010.

[70] Y. Liu, S. El-Naggar, D. S. Darling, Y. Higashi, and et al. Zeb1 links epithelial-mesenchymal transition and cellular senescence. *Development*, 135:579–588, 2008.

[71] W. Lo, L. Chen, M. Wang, and Q. Nie. A robust and efficient method for steady-state patterns in reaction-diffusion systems. *Jour. Comp. Physics*, 231:5062–5077, 2012.

[72] W. C. Lo, C. S. Chou, K. K. Gokoffski, and et al. Feedback regulation in multistage cell lineages. *Math Biosci Eng*, 6:59–82, 2009.

[73] G. J. Lord and J. Rougemont. A numerical scheme for stochastic PDEs with Gevrey regularity. *IMA Journal of Numerical Analysis*, 24:587–604, 2004.

[74] Gabriel J. Lord, Catherine E. Powell, and Tony Shardlow. *An introduction to computational stochastic PDEs*. Cambridge Texts in Applied Mathematics, 2014.

[75] M. Lu, M. K. Jolly, R. Gomoto, B. Huang, and et al. Tristability in cancer-associated microRNA-TF chimera toggle switch. *The Journal of Physical Chemistry B*, 117:13164–13174, 2013.

[76] M. Lu, M. K. Jolly, H. Levine, J. N. Onuchic, and E. Ben-Jacob. MicroRNA-based regulation of epithelial-hybrid-mesenchymal fate determination. *Proc. Natl. Acad. Sci. USA*, 110:18144–18149, 2013.

[77] P. K. Maini, T. E. Woolley, R. E. Baker, E. A. Gaffney, and S. S. Lee. Turing's model for biological pattern formation and the robustness problem. *Interface Focus*, 2:487–496, 2012.

[78] S. Mani, W. Guo, M. Liao, E. N. Eaton, A. Ayyanan, and et al. The epithelial-mesenchymal transition generates cells with properties of stem cells. *Cell*, 133(4):704–715, 2008.

[79] D. S. Micalizzi, S. M Farabaugh, and H. L. Ford. Epithelial-mesenchymal transition in cancer: parallels between normal development and tumor progression. *Journal of Mammary Gland Biol. Neoplasia*, 15:117–134, 2010.

[80] G. N. Milstein, E. Platen, and H. Schurz. Balanced implicit methods for stiff stochastic systems. *Siam J. Numer. Anal.*, 35(3):1010–1019, 1998.

[81] T. Müller-Gronbach and K. Ritter. An implicit Euler scheme with non-uniform time discretization for heat equations with multiplicative noise. *BIT Numer Math*, 47(2):393–418, 2007.

[82] Q. Nie, F. Wan, Y. Zhang, and X. Liu. Compact integration factor methods in high spatial dimensions. *Jour. Comp. Physics*, 227:5238–5255, 2008.

[83] Q. Nie, Y. Zhang, and R. Zhao. Efficient semi-implicit schemes for stiff systems. *Jour. Comp. Physics*, 214:521–537, 2006.

[84] M. A. Nieto. Epithelial plasticity: a common theme in embryonic and cancer cells. *Science*, 342(e1234850), 2013.

[85] M. A. Nieto, R. Huang, R. Jackson, and J. P. Thiery. EMT: 2016. *Cell*, 166(1):21–45, 2016.

[86] K-A. Norton, M. M. McCabe Pryor, and A. S. Popel. Multiscale modeling of cancer. *bioRxiv*, 033977, 2015.

[87] H. Peinado, M. Quintanilla, and A. Cano. Transforming growth factor beta-1 induces snail transcription factor in epithelial cell lines: mechanisms for epithelial mesenchymal transitions. *The Journal of Biological Chemistry*, 278:21113–21123, 2003.

[88] S. Peiró, M. Escrivà, I. Puig, and et al. Snail1 transcriptional repressor binds to its own promoter and controls its expression. *Nucleic Acids Research*, 34:2077–2084, 2006.

[89] K. V. Price, R. M. Storn, and J. A. Lampinen. *Differential Evolution : a Practical Approach to Global Optimization*. Springer, Berlin, 2005.

[90] I. Ramis-Conde, D. Drasdo, A. R. A. Anderson, and M. A. J. Chaplain. Modeling the influence of the E-Cadherin-β-Catenin pathway in cancer cell invasion: a multiscale approach. *Biophysical Journal*, 95:155–165, 2008.

[91] H. Roca, J. Hernandez, S. Weidner, R. C. McEachin, and et al. Transcription factors OVOL1 and OVOL2 induce the mesenchymal to epithelial transition in human cancer. *PLos One*, 8, 2013.

[92] Y. Shen, C. Shi, W. Wei, and et al. The heterogeneity and dynamic equilibrium of rat embryonic stem cells. *Cell Res*, 21:1143–1147, 2011.

[93] H. Siemens, R. Jackstadt, S. Hünten, M. Kaller, and et al. miR34 and SNAIL form a double-negative feedback loop to regulate epithelial-mesenchymal transitions. *Cell Cycle*, 10:4256–4271, 2011.

[94] M. S. Sosa, P. Bragado, and J. A. Aguirre-Ghiso. Mechanisms of disseminated cancer cell dormancy: an awakening field. *Nat Rev Cancer*, 14:611–622, 2014.

[95] Rainer Storn. Differential evolution (de) for continuous function optimization (an algorithm by kenneth price and rainer storn). `http://www1.icsi.berkeley.edu/~storn/code.html`. Accessed May 2016.

[96] R. Strauss, Z. Y. Li, Y. Liu, I. Beyer, J. Persson, P. Sova, T. Möller, S. Pesonen, and et al. Analysis of epithelial and mesenchymal markers in ovarian cancer reveals phenotypic heterogeneity and plasticity. *PLos One*, 6(1):e16186, 2011.

[97] C. Ta, Q. Nie, and T. Hong. Controlling stochasticity in epithelial-mesenchymal transition through multiple intermediate cellular states. *Dis. Cont. Dynamical Systems - Series B*, 21(7):2275–2291, 2016.

[98] S. Di Talia, J. M. Skotheim, J. M. Bean, and et al. The effects of molecular noise and size control on variability in the budding yeast cell cycle. *Nature*, 448:947–951, 2007.

[99] W. L. Tam and R. A. Weinberg. The epigenetics of epithelial-mesenchymal plasticity in cancer. *Nat. Med*, 19:14381449, 2013.

[100] X. J. Tian, H. Zhang, and J. Xing. Coupled reversible and irreversible bistable switches underlying TGFβ-induced epithelial to mesenchymal transition. *Biophysical Journal*, 105(4):1079–1089, 2013.

[101] J. J. Tyson and B. Novak. Functional motifs in biochemical reaction networks. *Annu Rev Phys Chem*, 61:219–240, 2010.

[102] A. Villarejo, A. Cabrera-Cort'es, P. Molina-Ort'iz, and et al. Differential role of Snail1 and Snail2 zinc fingers in E-cadherin repression and epithelial to mesenchymal transition. *The Journal of Biological Chemistry*, 289:930–941, 2014.

[103] B. Wang, P. Koh, C. Winbanks, M.T. Coughlan, and et al. miR-200a prevents renal fibrogenesis through repression of TGF-β2 expression. *Diabetes*, 60:280–287, 2011.

[104] D. Wang, L. Zhang, and Q. Nie. Array-representation integration factor method for high-dimensional systems. *Jour. Comp. Physics*, 258:585–600, 2014.

[105] L. Wang, J. Xin, and Q. Nie. A critical quantity for noise attenuation in feedback systems. *PLoS Comput Biol*, 6, 2010.

[106] Liming Wang, Jack Xin, and Qing Nie. A critical quantity for noise attenuation in feedback systems. *PLoS Comput Biol*, 6:e1000764, 04 2010.

[107] X. Wang, S. Gan, and D. Wang. A family of fully implicit Milstein methods for stiff stochastic differential equations with multiplicative noise. *BIT Numer Math*, 52(3):741–772, 2012.

[108] K. Watanabe, A. Villarreal-Ponce, P. Sun, M. L. Salmans, and et al. Mammary morphogenesis and regeneration require the inhibition of EMT at terminal end buds by Ovol2 transcriptional repressor. *Dev Cell*, 29:59–74, 2014.

[109] W. Weston, R. Perez J. Zayas, and et al. Dynamic equilibrium of heterogeneous and interconvertible multipotent hematopoietic cell subsets. *Sci Rep*, 4, 2014.

[110] D. J. Wilkinson. Stochastic modeling for quantitative description of heterogeneous biological systems. *Nat Rev Genet*, 10(1471-0056):122–133, 2009.

[111] K. Williams, K. Motiani, P. V. Giridhar, and S. Kasper. CD44 integrates signaling in normal stem cell, cancer stem cell and (pre)metastatic niches. *Exp. Biol. Med. (Maywood)*, 238:324–338, 2013.

[112] I. Wong, S. Javaid, E. Wong, S. Perk, and el al. Collective and individual migration following the epithelial-mesenchymal transition. *Nature Materials*, 13, 2014.

[113] X. Ye, W. L. Tam, T. Shibue, and et al. Distinct EMT programs control normal mammary stem cells and tumour-initiating cells. *Nature*, 525:256–260, 2015.

[114] M. Yu, A. Bardia, B. S. Wittner, S. L. Stott, and et al. Circulating breast tumor cells exhibit dynamic changes in epithelial and mesenchymal composition. *Science*, 339:580–584, 2013.

[115] J. Zhang, X. J. Tian, H. Zhang, Y. Teng, R. Li, and et al. TGFβ-induced epithelial to mesenchymal transition proceeds through stepwise activation of multiple feedback loops. *Science Signaling*, 7(345), 2014.

[116] L. Zhang, K. Radtke, L. Zheng, A. Q. Cai, T. F. Schilling, and Q. Nie. Noise drives sharpening of gene expression boundaries in the zebrafish hindbrain. *Molecular Systems Biology*, 8(613), 2012.

[117] T. Zhang, Q. Zhu, Z. Xie, Y. Chen, and et al. The zinc finger transcription factor Ovol2 acts downstream of the bone morphogenetic protein pathway to regulate the cell fate decision between neuroectoderm and mesendoderm. *Journal of Biol. Chem.*, 288:6166–6177, 2013.

[118] L. Zheng, M. Chen, and Q. Nie. External noise control in inherently stochastic biological systems. *J Math Phys*, 53, 2012.

[119] J. X. Zhou, A. O. Pisco, H. Qian, and S. Huang. Nonequilibrium population dynamics of phenotype conversion of cancer cells. *PLoS One*, 9:e110714e110714, 2014.

Appendices

A Experimental evidence confirms the existence of an Ovol2-Zeb1 mutual inhibition loop (Chapter 2)

Mutual inhibition loops between EMT-inducing TFs and miRNAs (e.g. Zeb1-miR200 and Snail-miR34a) are critical for robust control of EMT/MET [84]. Previous studies showed that Zeb1 is directly inhibited by Ovol2 in mammary and skin epithelial cells [61, 108]. Zeb1 and Ovol2 are expressed in a mutually exclusive pattern in clinical and cell line samples [91, 108], raising the possibility that Zeb1 may also inhibit Ovol2 expression. Indeed, sequence analysis revealed the presence of two conserved Zeb1 binding consensus sequences in the human and mouse *OVOL2*/Ovol2 promoters, one near the transcriptional start site (-335 bp and -111 bp for human and mouse genes, respectively) and the other further upstream (-1546 bp and -1167 bp for human and mouse, respectively), Figure H.1A. Using chromatin immunoprecipitation (ChIP) assay, Xing Dai's lab detected Zeb1 binding to the downstream but not upstream site, Figure H.1A. Furthermore, forced expression of Zeb1 in MCF10A human mammary epithelial cells significantly decreased *OVOL2* expression at a transcriptional level, whereas Ovol2 overexpression led to reduced level of ZEB1 transcript as expected, Figure H.1B. These results are consistent with direct repression of *OVOL2*/Ovol2 expression by Zeb1, and together with previously published data suggest the existence of an Ovol2-Zeb1 mutual

inhibition loop.

B Experimental evidence supports the 4-state model and validates the reprogramming ability of Ovol2 (Chapter 2)

Previous work stipulates that unstimulated MCF10A cells are in an epithelial state and when stimulated by increasing concentrations of TGFβ they transition into first an intermediate (partial EMT) state and subsequently an M state [115]. However, by comparing the expression of epithelial (E-cadherin or Ecad) and mesenchymal (Vimentin or Vim) markers between MCF10A and two breast cancer cell lines well-characterized for their cellular states (MCF7 = E state, MDA-MB231 = metastatic human breast cancer cells corresponding to an M state), we found MCF10A cells to be likely in a state that is intermediate between typical terminal E and M cells (Figure 2.2A, compare green population to others; Figure H.2). This is consistent with a recent study showing that MCF10A cells tend to collectively migrate [112], a feature that has been associated with the intermediate phenotype [76, 79, 114]. We surmise that the natural state of these cells is I_1, because a majority of them show low to no Vim expression, suggesting more similarity to the terminal E than M cells.

To experimentally test whether Ovol2 is able to reprogram I_1-state cells into an E state (Figure 2.1C), Xing Dai's lab overexpressed Ovol2 in MCF10A cells using a lentiviral expression system in which the Ovol2-expressing cells can be distinguished from uninfected cells by bicistronic expression of GFP, Figure H.3. This led to significantly increased expression of Ecad, and decreased expression of Vim as assessed by quantitative population analysis using flow cytometry (Figure 2.2B, blue population). Comparison with Ecad/Vim profiles in Figure 2.2A reveals the similarity between Ovol2-reprogrammed cells and terminal E cells

(MCF7). In contrast, overexpression of EMT inducers Snail or Zeb1 directed MCF10A cells to an M phenotype (Figure 2.2B, red and orange populations). Consistent with previous reports that EMT promotes stemness [12, 31, 78], the expression of a well-known cancer stem cell marker CD44 [111] decreased upon Ovol2-induced transition to E and increased upon Zeb1/Snail-induced transition to M, Figure 2.2D.

As bifurcation analysis also predicted the ability of Ovol2 to reprogram M-state cells into an E state (Figure 2.1C), we tested the effect of Ovol2 overexpression on MDA-MB231 cells. Indeed, forced expression of Ovol2 was able to convert these cells to exhibiting a pattern of Ecad/Vim expression that is reminiscent of the terminal E state, Figure 2.2E. This finding is consistent with previous reports of Ovol2 overexpression inducing epithelial features in M-state cells [91, 108].

C Model construction (Chapter 2)

The framework of our model is from a recent publication by Zhang et al [115]. In that study, Zhang et al. demonstrated that both Zeb1-miR-200 and Snail1-miR-34 loops can create bistable switches. Therefore, this published model is also referred to as Cascading Bistable Switches (CBS) model in that paper and herein. We added the following interactions to the existing network in CBS model in order to consider regulations involving Ovol2:

1. Ovol2 inhibits Zeb1 expression [61, 108]
2. Ovol2 inhibits TGFβ expression [108]
3. Ovol2 inhibits TGFβ signaling, partly by inhibiting Snail expression [87, 108]
4. Zeb1 inhibits Ovol2 expression (evidence from this study)

Other regulations and the supporting experimental evidence are listed in Table G.1.

Details of modeling framework can be found in [115]. Briefly, law of mass action was used to describe miRNA-mRNA interaction, and miRNA-mediated mRNA degradation was ex-

plicitly considered. This framework was first introduced by Lu et al [75, 76]. Generalized Hill function was used to describe transcriptional activation and inhibition by transcription regulators other than miRNAs. We adjusted some parameter values from the CBS model in order to examine the qualitative dynamic behavior that might be introduced by Ovol2-Zeb1 interactions.

Our model has 18 ordinary differential equations (ODEs) as follows:

$$\frac{d[snail1]_t}{dt} = k0_{snail} +$$
$$k_{snail}\left(\frac{(([TGF]+[TGF0])/J0_{snail})^{n0_{snail}}}{(1+(([TGF]+[TGF0])/J0_{snail})^{n0_{snail}}+([OVOL2]/J1_{snail})^{n1_{snail}})}\right)$$
$$\left(\frac{1}{1+([SNAIL]/J2_{snail})}\right) - kd_{snail}([snail1]_t - [SR]) - kd_{SR}[SR]$$
$$\frac{d[SNAIL]}{dt} = k_{SNAIL}([snail1]_t - [SR]) - kd_{SNAIL}[SNAIL]$$
$$\frac{d[miR34]_t}{dt} = k0_{34} + \frac{k_{34}}{1+([SNAIL]/J1_{34})^{n1_{34}}+([ZEB]/J2_{34})^{n2_{34}}}$$
$$- kd_{34}([miR34]_t - [SR]) - (1-\lambda_{SR})kd_{SR}[SR]$$
$$\frac{d[SR]}{dt} = Tk(K_{SR}([snail1]_t - [SR])([miR34]_t - [SR]) - [SR])$$
$$\frac{d[zeb]_t}{dt} = k0_{zeb} + k_{zeb}\frac{([SNAIL]/J1_{zeb})^{n1_{zeb}}}{1+([SNAIL]/J1_{zeb})^{n1_{zeb}}+([OVOL2]/J2_{zeb})^{n2_{zeb}}} - kd_{[zeb]}([zeb]_t$$
$$- \sum_{i=1}^{5} C_5^i[ZR_i]) - \sum_{i=1}^{5} kd_{ZR_i}C_5^i[ZR_i]$$
$$\frac{d[ZEB]}{dt} = k_{ZEB}([zeb]_t - \sum_{i=1}^{5} C_5^i[ZR_i]) - kd_{ZEB}[ZEB]$$

$$\frac{d[miR200]_t}{dt} = k0_{200} + \frac{k_{200}}{1 + ([SNAIL]/J1_{200})^{n1_{200}} + ([ZEB]/J2_{200})^{n2_{200}}}$$

$$- kd_{200}([miR200]_t - \sum_{i=1}^{5} iC_5^i[ZR_i] - [TR]) - \sum_{i=1}^{5}(1-\lambda_i)kd_{ZR_i}C_5^i i[ZR_i]$$

$$- (1-\lambda_{TR})kd_{TR}[TR]$$

$$\frac{d[ZR_1]}{dt} = Tk(K_1([miR200]_t - \sum_{i=1}^{5} iC_5^i[ZR_i] - [TR])([zeb]_t - \sum_{i=1}^{5} C_5^i[ZR_i]) - [ZR_1])$$

$$\frac{d[ZR_2]}{dt} = Tk(K_2([miR200]_t - \sum_{i=1}^{5} iC_5^i[ZR_i] - [TR])[ZR_1] - [ZR_2])$$

$$\frac{d[ZR_3]}{dt} = Tk(K_3([miR200]_t - \sum_{i=1}^{5} iC_5^i[ZR_i] - [TR])[ZR_2] - [ZR_3])$$

$$\frac{d[ZR_4]}{dt} = Tk(K_4([miR200]_t - \sum_{i=1}^{5} iC_5^i[ZR_i] - [TR])[ZR_3] - [ZR_4])$$

$$\frac{d[ZR_5]}{dt} = Tk(K_5([miR200]_t - \sum_{i=1}^{5} iC_5^i[ZR_i] - [TR])[ZR_4] - [ZR_5])$$

$$\frac{d[tgf]_t}{dt} = k_{tgf} - kd_{tgf}([tgf]_t - [TR]) - kd_{TR}[TR]$$

$$\frac{d[TGF]}{dt} = k0_{TGF} + k_{TGF}([tgf]_t - [TR]) - kd_{TGF}[TGF]$$

$$\frac{d[TR]}{dt} = Tk(K_{TR}([miR200]_t - \sum_{i=1}^{5} iC_5^i[ZR_i] - [TR])([tgf]_t - [TR]) - [TR])$$

$$\frac{d[Ecad]}{dt} = k0_E + \frac{k_{E1}}{1 + ([SNAIL]/J1_E)^{n1_E}} + \frac{k_{E2}}{1 + ([ZEB]/J2_E)^{n2_E}} - kd_E[Ecad]$$

$$\frac{d[Vim]}{dt} = k0_V + \left(k_{V1}\frac{([SNAIL]/J1_V)^{n1_V}}{1 + ([SNAIL]/J1_V)^{n1_V}} + k_{V2}\frac{([ZEB]/J2_V)^{n2_V}}{1 + ([ZEB]/J2_V)^{n2_V}}\right)$$

$$/(1 + [OVOL2]/J3_V) - kd_V[Vim]$$

$$\frac{d[OVOL2]}{dt} = k0_O + \frac{k_O}{1 + ([ZEB]/J_O)^{n_O}} - kd_O[OVOL2]$$

where $\sum_{i=1}^{5} iC_5^i[ZR_i] = 5[ZR_1] + 2*10[ZR_2] + 3*10[ZR_3] + 4*5[ZR_4] + 5[ZR_5]$, which represents the total amount of miR-200 bound to *Zeb1* mRNA. Each term of this summation describes C_5^i scenarios in which i number of miRNA molecules bind to C_5^i possible binding sites that each *Zeb1* mRNA might have. $\sum_{i=1}^{5} C_5^i[ZR_i] = 5[ZR_1] + 10[ZR_2] + 10[ZR_3] +$

$5[ZR_4]+[ZR_5]$ represents the total amount of complex formed by miR-200 and *Zeb1* mRNA.

Meanings of the state variables are as follows:

$[snail1]_t$: Concentration of *Snail1* mRNA

$[SNAIL]$: Concentration of Snail protein

$[miR34]_t$: Concentration of miR-34

$[SR]$: Concentration of complex formed by miR-34 and *Snail1* mRNA

$[zeb]_t$: Concentration of *Zeb1* mRNA

$[ZEB]$: Concentration of Zeb1 protein

$[miR200]_t$: Concentration of miR-200

$[ZR_i]$: Concentrations of complexes formed by i miR-200 molecules and one *Zeb1* mRNA molecule

$[tgf]_t$: Concentration of $TGF\beta$ mRNA

$[TGF]$: Concentration of endogenous TGFβ protein

$[TR]$: Concentration of complex formed by miR-34 and $TGF\beta$ mRNA

$[Ecad]$: Concentration of Ecad protein

$[Vim]$: Concentration of Vim protein

$[OVOL2]$: Concentration of Ovol2 protein

D Stochastic simulations (Chapter 2)

To consider the effect of molecular fluctuations in gene regulatory networks, we introduced uncorrelated multiplicative white noise to some of the ODEs as $dx_i/dt = f_i(x_1, x_2, ..., x_n) + \mu_{x_i} x_i dW_i$, $i = 1, ..., n$, where dW_i is a Wiener process that can be discretized as $dW_i = \sqrt{dt}\mathcal{N}(0,1)$ where $\mathcal{N}(0,1)$ denotes a normally distributed random variable with 0 mean and unit variance, and μ_{x_i} represents the amplitude of the fluctuation. We noticed that the fluctu-

ations on EMT -inducing and -repressing factors may have different effects on noise-induced phenotypic transitions (not shown), so we introduced the white noise to ODEs for Snail and Ovol2 to represent the noise on these two groups of factors. To adapt to the heterogeneous populations that we observed in flow cytometry experiments and to consider the additional fluctuations due to detection inaccuracy, we also introduced white noise to ODEs for Ecad and Vim. The basal values of the amplitude parameters for these four stochastic ODEs are $\mu_{Snail1} = 0.02$, $\mu_O = 0.2$, $\mu_E = 0.02$, and $\mu_V = 0.02$. To compare the effects of different levels of fluctuations, we increased μ_{Snail1} and/or μ_O by 10-fold in some of our analysis (Figures 2.2, H.3, H.5). Larger fluctuations for Ecad and Vim ($\mu_E = 0.07$ and $\mu_V = 0.07$) were used to adapt to the experimental observations from flow cytometry (Figure 2.2).

To unambiguously determine the phenotypes by the end of the stochastic simulations ($t = 500$), we set the amplitude parameters μ_{x_i} to 0 at $t = 500$ and continued the simulations for another 100 time units so that each trajectory is settled at a deterministic steady state (Figure H.5).

The high complexity of the system prevents us from more formal implementation of stochastic simulations such as Gillespie algorithms. A more systematic study of the effects of fluctuations and stochastic transitions among multiple states will be a subject of future studies.

E Analysis of the effects of specific regulations on cell phenotypes (Chapter 2)

One-parameter bifurcation analysis was used to examine the change of the stable cell phenotypes in response to change of a particular parameter (e.g. external TGFβ concentration or basal production rate of Ovol2). Two-parameter bifurcation analysis was used to examine the phenotypic changes in response to both EMT -inducing and -repressing parameters/signals. The curves on the two-parameter bifurcation diagrams are continuation of the saddle-node

bifurcation points that we obtained from the one-parameter bifurcation analysis (Figure 2.4). To simulate the overexpression experiment, we increased in basal production rates of Ovol2 or Zeb1 by 10-fold (Figures 2.2 and H.3). To understand how each regulation of the network contributes to overall performance of the system, we used 'knockdown' approach in which we partially or completely removed certain regulations in the network and examined the effect on the two-parameter bifurcation diagrams (see Figure 2.6). To simulate partial removal of Zeb1-Ovol2 mutual inhibition, we increased the Michaelis constants of the inhibitions between these two molecules (J_O and $J2_{zeb}$) by 2-fold. For complete removal, we increased these parameters by 100-fold. Similar approaches were used to simulate complete removal of TF-microRNA mutual inhibition loops, i.e. $J1_{34}$ and/or $J2_{200}$ were increased by 100-fold, whereas K_1 and K_{SR} (TF-microRNA binding/unbinding rate ratios) were decreased by 100-fold. Although it is challenging to systematically demonstrate the global effect of a particular regulation in a large parameter space, we assume that this approach gives a reasonable representation on their general functions. We did not analyze functions of isolated 'modules' in the network since they have been discussed extensively in previous computational analysis of the EMT system [75, 76, 100, 115]. Also, combination of mutual-inhibition and auto-activation loops is a well-studied dynamic system in general, so detailed analysis on Zeb1-Ovol2 module is not necessary.

F Methods (Chapter 2)

F.1 Cell lines

MCF10A, MCF7, MDA-MB-231 were purchased from ATCC. MCF7 and MDA-MB231 cells were maintained in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% fetal bovine serum. MCF10A cells were grown in DMEM/F12 (1:1) medium with 5% horse

serum, epidermal growth factor (10 ng/mL), cholera toxin (100 ng/mL) and insulin (0.023 IU/mL). For TGFβ treatment, cells were incubated with titrated concentrations of human TGFβ1 protein (R & D systems) in complete culture medium for 10 days. The culture medium was replaced daily and cells were passaged just before reaching full confluency.

F.2 Lentiviral expression system

Recombinant lentiviruses expressing Ovol2 using the pHIV-ZsGreen lentiviral construct was described previously [108]. For Snail and Zeb1 expression, human SNAIL cDNA and mouse Zeb1 cDNA were cloned into the XhoI/NotI and EcoRI sites of pHIV-ZsGreen, respectively. Production and infection of lentiviruses were carried out as previously described [108]. Transduction unit of viral solution was estimated by measuring GFP-positive population using a flow cytometer.

F.3 Flow cytometry

For Ecad/Vim profiling, cells were fixed with 4% paraformaldehyde, permeabilized with 0.2% Triton X-100, and stained with the following primary and secondary antibodies and reagents: anti-E-cadherin (Life Technologies, 1:500), anti-vimentin (Cell Signaling Technologies, 1:500), allophycocyanin (APC)-labeled anti-mouse IgG (Santa Cruz), Cy3-labeled anti-rabbit IgG (Jackson Immuno). For CD44 staining, live cells were stained with phycoerythrin (PE)-conjugated anti-CD44 antibody (Biolegend). Live-cell sorting for GFP+ cells was performed on a BD FACSAria equipped with FACS DiVa6.0 software operating at low pressure (20 psi) using a 100-μm nozzle. Cell clusters and doublets were electronically gated out. Cells were routinely double sorted and post-sort analysis typically indicated purities of >90% with minimal cell death (<10%).

F.4 ChIP assay

ChIP was performed with an anti-Zeb1 antibody (Santa Cruz) according to the previously described protocol [108]. The following primers were used to detect Ovol2 promoter regions: proximal site (forward, 5'-GTGATAGGGGTATGAAGCAGAGG-3' reverse, 5'-CACCAGGAAACTTGGGAGTG-3') and distal site (forward, 5'-AGCCCAGAAATCCGTTACCA-3' reverse, 5'-CTCACTGCTGGAGGTTGTCT-3').

F.5 RT-PCR

Total RNA was isolated using the TRIzol Reagent (Invitrogen) followed by cleaning up and RNase-free DNaseI treatment using the RNeasy mini kit (QIAGEN). cDNA was prepared using Retroscript Kit (Applied Biosystems) according to manufacturer's instructions. Realtime PCR was performed using a CFX96 qPCR system and SsoAdvanced SYBR Green Supermix (Bio-rad). Comparative analysis was performed between the genes of interest normalized by the house keeping genes *GAPDH* and *ACTB*. The following primers were used: *OVOL2* (forward, 5'-AGCTGTGACCTGTGTGGCAAG-3' reverse, 5'-ACGAATGCCTGTGTGTGTGC-3'), *ZEB1* (forward 5'-TTGCTCCCTGTGCAGTTACA-3' reverse 5'-CGTTTCTTGCAGTTTGGGCA-3'), *GAPDH* (forward, 5'-GGACCTGACCTGCCGTCTAGAA-3' reverse, 5'-GGTGTCGCTGTTGAAGTCAGAG-3'), and *ACTB* (forward, 5'-CTTCTACAATGAGCTGCGTG-3' reverse, 5'-GGGTGTTGAAGGTCTCAAAC-3'). For semi-quantitative PCR, the following primers were used: *CDH1* (forward, 5'-AAAGGCCCATTTCCTAAAAACCT-3' reverse, 5'-TGCGTTCTCTATCCAGAGGCT-3'), *SNAI1* (forward, 5'-CCTCCCTGTCAGATGAGGAC-3' reverse, 5'-CCAGGCTGAGGTATTCCTTG-3'), *VIM* (forward, 5'-GACGCCATCAACACCGAGTT-3' reverse, 5'-CTTTGTCGTTGGTTAGCTGGT-3').

F.6 Mathematical modeling

We used ordinary differential equations (ODEs) to model the regulatory network shown in Figure 2.1A. The framework of the model stems from a recently published EMT model [115], and the modeling details are described therein. This framework employs mass-action dynamics to model microRNA-mRNA interactions with considerations of the microRNA binding sites on their targets. This modeling strategy was introduced by Lu et al [75, 76]. As other transcription factor regulations, interactions involving Ovol2 were modeled with Hill functions. Numerical bifurcation analysis was performed with PyDSTool [17]. To consider fluctuations in gene expression, we added multiplicative white noise to some of the ODEs. To determine which phenotype (basin of attraction) a cell adopts at the end of the simulations, we set the noise terms to zero and let the simulation continue until it reached steady state (Figure H.7). Lists of equations, parameters and assumptions can be found in supplementary materials. Stochastic simulations were performed with XPPAUT [22].

G Supplementary tables (Chapter 2)

Table G.2: List of basal parameter values

Name	Description	Value
$J1_{200}$	Michaelis constant of Snail-dependent inhibition of miR-200 production	3μM
$J2_{200}$	Michaelis constant of Zeb1-dependent inhibition of miR-200 production	0.2μM
$J1_{34}$	Michaelis constant of Snail-dependent inhibition of miR-34 production	0.15μM
Continued on next page		

Table G.2 – continued from previous page

Name	Description	Value
$J2_{34}$	Michaelis constant of Zeb1-dependent inhibition of miR-34 production	0.35μM
J_O	Michaelis constant of Zeb1-dependent inhibition of Ovol2 production	0.9μM
$J0_{snail}$	Michaelis constant of TGFβ-dependent $Snail1$ mRNA transcription	0.6μM
$J1_{snail}$	Michaelis constant of Ovol2-dependent inhibition of $Snail1$ transcription	0.5μM
$J2_{snail}$	Michaelis constant of Snail self-inhibition	1.8μM
$J1_E$	Michaelis constant of Snail-dependent inhibition of Ecad production	0.1μM
$J2_E$	Michaelis constant of Zeb1-dependent inhibition of Ecad production	0.3μM
$J1_V$	Michaelis constant of Snail-dependent Vim production	0.4μM
$J2_V$	Michaelis constant of Zeb1-dependent Vim production	0.4μM
$J3_V$	Michaelis constant of Ovol2-dependent inhibition of Vim production	2μM
$J1_{zeb}$	Michaelis constant of Snail-dependent $Zeb1$ transcription	3.5μM
$J2_{zeb}$	Michaelis constant of Ovol2-dependent inhibition of $Zeb1$ transcription	0.9μM
K_1	Binding/unbinding rate ratio for the first miR-200 molecule and $Zeb1$ mRNA	$1.0/\mu$M
K_2	Binding/unbinding rate ratio for the second miR-200	$1.0/\mu$M

Continued on next page

Table G.2 – continued from previous page

Name	Description	Value
	molecule and $Zeb1$ mRNA	
K_3	Binding/unbinding rate ratio for the third miR-200 molecule and $Zeb1$ mRNA	$1.0/\mu$M
K_4	Binding/unbinding rate ratio for the fourth miR-200 molecule and $Zeb1$ mRNA	$1.0/\mu$M
K_5	Binding/unbinding rate ratio for the fifth miR-200 molecule and $Zeb1$ mRNA	$1.0/\mu$M
K_{TR}	Binding/unbinding rate ratio for the first miR-200 molecule and $TGF\beta$ mRNA	$20/\mu$M
K_{SR}	Binding/unbinding rate ratio for the first miR-34 molecule and $Snail1$ mRNA	$100/\mu$M
$TGF0$	Concentration of exogenous TGFβ	0μM
Tk	Timescale constant for all miRNA-mRNA binding/unbinding events	1000
$k0_O$	Basal production rate of Ovol2	0.35μM/hr
$k0_{200}$	Basal production rate of miR-200	0.0002μM/hr
$k0_{34}$	Basal production rate of miR-34	0.001μM/hr
$k0_{snail}$	Basal production rate of $Snail1$ mRNA	0.0005μM/hr
$k0_{zeb}$	Basal production rate of $Zeb1$ mRNA	0.003μM/hr
$k0_{TGF}$	Basal production rate of TGFβ	1.1μM/hr
$k0_E$	Basal production rate of Ecad	5μM/hr
$k0_V$	Basal production rate of Vim	5μM/hr
k_{E1}	Production rate 1 of Ecad	15μM/hr

Continued on next page

Table G.2 – continued from previous page

Name	Description	Value
k_{E2}	Production rate 2 of Ecad	5μM/hr
k_{V1}	Production rate 1 of Vim	2μM/hr
k_{V2}	Production rate 2 of Vim	5μM/hr
k_O	Production rate of Ovol2	1.2μM/hr
k_{200}	Production rate of miR-200	0.02μM/hr
k_{34}	Production rate of miR-34	0.01μM/hr
k_{snail}	Transcription rate of *Snail*1 mRNA	0.05μM/hr
k_{tgf}	Transcription rate of TGFβ	0.05μM/hr
k_{zeb}	Transcription rate of *Zeb*1 mRNA	0.06 μM/hr
k_{TGF}	Transcription rate of TGFβ protein	1.5μM/hr
k_{SNAIL}	Transcription rate of Snail protein	16μM/hr
k_{ZEB}	Transcription rate of Zeb1 protein	16μM/hr
kd_{ZR_1}	Degradation rate for complex of 1 *Zeb*1 mRNA and 1 miR-200 molecules	0.5/hr
kd_{ZR_2}	Degradation rate for complex of 1 *Zeb*1 mRNA and 2 miR-200 molecules	0.5/hr
kd_{ZR_3}	Degradation rate for complex of 1 *Zeb*1 mRNA and 3 miR-200 molecules	0.5/hr
kd_{ZR_4}	Degradation rate for complex of 1 *Zeb*1 mRNA and 4 miR-200 molecules	0.5/hr
kd_{ZR_5}	Degradation rate for complex of 1 *Zeb*1 mRNA and 5 miR-200 molecules	0.5/hr
kd_O	Degradation rate of Ovol2	1.0/hr

Continued on next page

Table G.2 – continued from previous page

Name	Description	Value
kd_{200}	Degradation rate of miR-200	0.035/hr
kd_{34}	Degradation rate of miR-34	0.035/hr
kd_{SR}	Degradation rate of complex *Snail*1 mRNA and miR-34	0.9/hr
kd_E	Degradation rate of Ecad	0.05/hr
kd_V	Degradation rate of Vim	0.05/hr
kd_{snail}	Degradation rate of complex *Snail*1 mRNA	0.09/hr
kd_{tgf}	Degradation rate of $TGF\beta$ mRNA	0.1/hr
kd_{TR}	Degradation rate of complex $TGF\beta$ mRNA and miR-200	1.0/hr
kd_{zeb}	Degradation rate of *Zeb*1 mRNA	0.10/hr
kd_{SNAIL}	Degradation rate of Snail protein	1.6/hr
kd_{TGF}	Degradation rate of TGFβ protein	0.9/hr
kd_{ZEB}	Degradation rate of Zeb1 protein	1.66/hr
$n1_{200}$	Hill exponent of Snail-dependent inhibition of miR-200 production	3
$n2_{200}$	Hill exponent of Zeb1-dependent inhibition of miR-200 production	2
$n1_{34}$	Hill exponent of Snail-dependent inhibition of miR-34 production	2
$n2_{34}$	Hill exponent of Zeb1-dependent inhibition of miR-34 production	2
n_O	Hill exponent of Zeb1-dependent inhibition of Ovol2 production	2
$n0_{snail}$	Hill exponent of TGFβ-dependent *Snail*1 mRNA transcription	2
$n1_{snail}$	Hill exponent of Ovol2-dependent inhibition of *Snail*1	2

Continued on next page

Table G.2 – continued from previous page

Name	Description	Value
	mRNA transcription	
$n1_E$	Hill exponent of Snail-dependent inhibition of Ecad production	2
$n2_E$	Hill exponent of Zeb1-dependent inhibition of Ecad production	2
$n1_V$	Hill exponent of Snail-dependent Vim production	2
$n2_V$	Hill exponent of Zeb1-dependent Vim production	2
$n1_{zeb}$	Hill exponent of Snail-dependent $Zeb1$ transcription	2
$n2_{zeb}$	Hill exponent of Ovol2-dependent inhibition of $Zeb1$ transcription	6
λ_1	Recycle rate of miR-200 during degradation of complex of 1 $Zeb1$ mRNA and 1 miR-200 molecules	0.5
λ_2	Recycle rate of miR-200 during degradation of complex of 1 $Zeb1$ mRNA and 2 miR-200 molecules	0.5
λ_3	Recycle rate of miR-200 during degradation of complex of 1 $Zeb1$ mRNA and 3 miR-200 molecules	0.5
λ_4	Recycle rate of miR-200 during degradation of complex of 1 $Zeb1$ mRNA and 4 miR-200 molecules	0.5
λ_5	Recycle rate of miR-200 during degradation of complex of 1 $Zeb1$ mRNA and 5 miR-200 molecules	0.5
λ_{SR}	Recycle rate of miR-34 during degradation of complex $Snail$ mRNA and miR-34	0.5
λ_{TR}	Recycle rate of miR-200 during degradation of complex $TGF\beta$ mRNA and miR-200	0.8

Molecular influence	*Evidence*
TGFβ upregulates *Snail*1 transcription	[87]
Snail upregulates *Zeb*1 transcription	[30]
Snail downregulates miR-34 transcription	[53, 93]
miR-34 downregulates Snail production	[53, 93]
Snail downregulates miR-200 transcription	[20]
Snail downregulates its own transcription	[88]
Snail downregulates Ecad production	[11, 102]
Snail upregulates Vim production	[102]
Zeb1 downregulates miR-200 transcription	[6, 7]
miR-200 downregulates Zeb1 production	[6, 7]
Zeb1 downregulates miR-34 transcription	[2]
Zeb1 downregulates Ecad production	[21]
Zeb1 upregulates Vim production	[70]
miR-200 downregulates TGFβ production	[28, 103]
Ovol2 downregulates TGFβ3 transcription	[108]
Ovol2 downregulates TGFβ signaling	[108]
Ovol2 downregulates Zeb1 transcription	[61, 108]
Zeb1 downregulates Ovol2 production	[40, 91]
Ovol2 downregulates Vim production	[61, 108]

Table G.1: Experimental evidence supporting influence diagram

H Supplementary figures (Chapter 2)

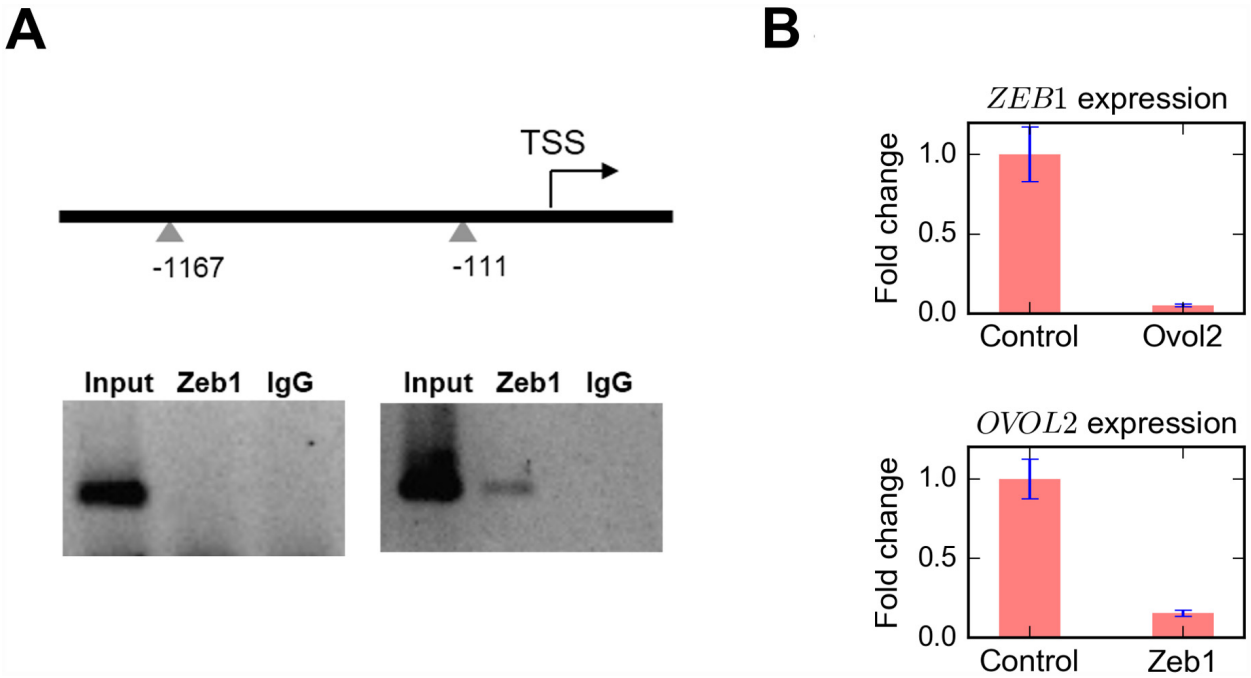

Figure H.1: Experimental evidence for Ovol2-Zeb1 mutual repression. **A)** Zeb1 binds to the *Ovol2* promoter in mouse mammary epithelial cells. Top, diagram showing the presence of Zeb1 consensus motifs (triangles) in *Ovol2* promoter. Bottom, ChIP-PCR using primers for the upstream (left) and downstream sites (right). **B)** RT-quantitative PCR analysis showing that OVOL2 and ZEB1 overexpression in MCF10A cells results in decreased level of *Zeb1* and *Ovol2* transcripts, respectively TSS, transcription start site.

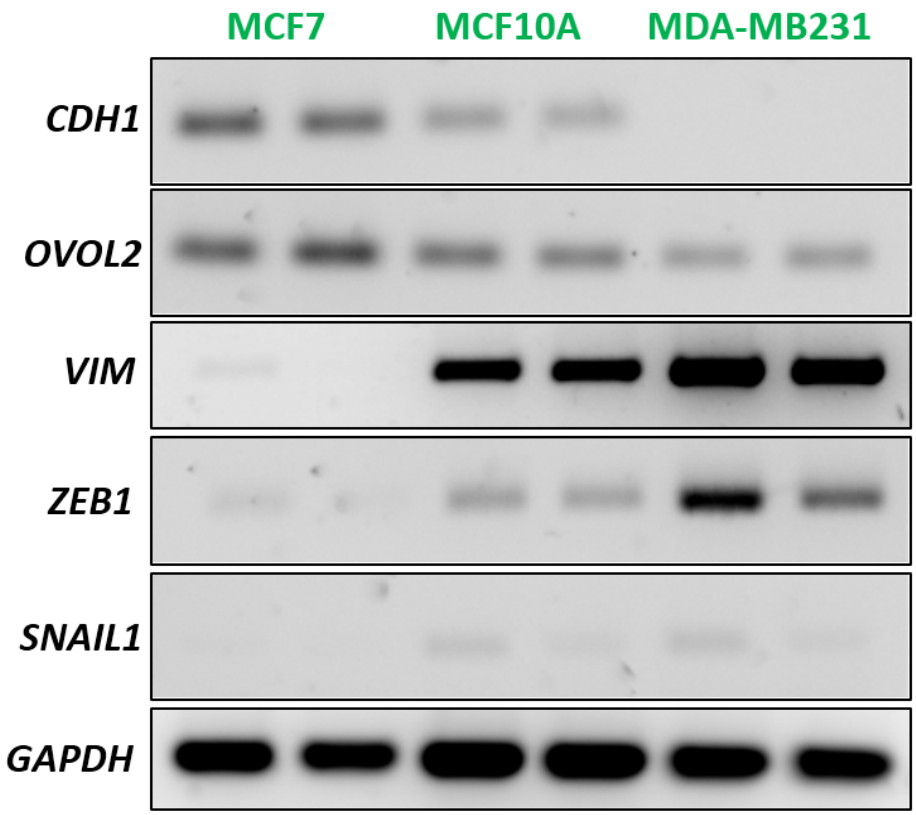

Figure H.2: MCF10A cells show intermediate expression levels of EMT-related genes. RT-semi-quantitative PCR analysis of the indicated genes in three human breast cell lines. Two biological replicates were performed for each gene in each type of cell line.

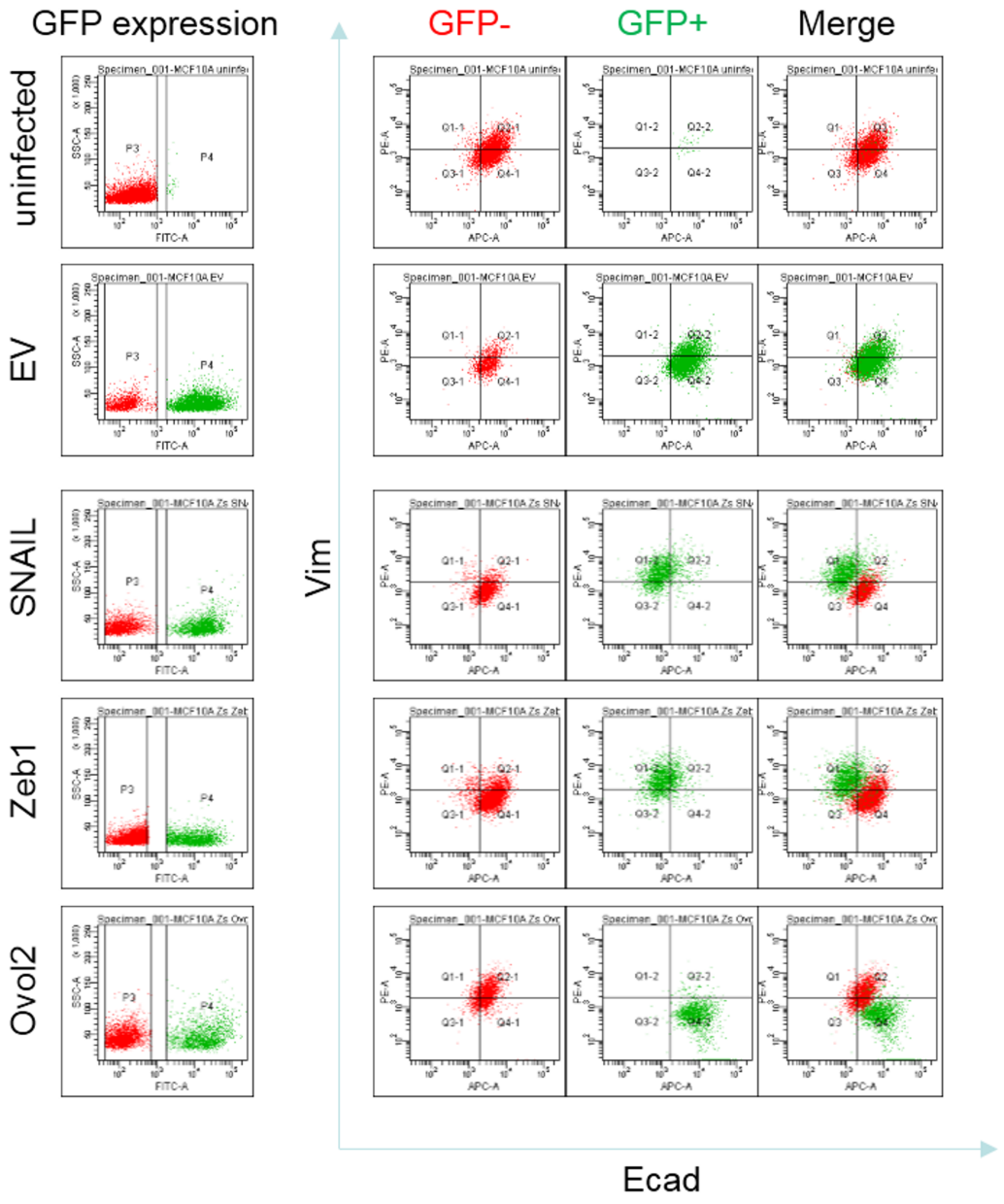

Figure H.3: FACS profiles of epithelial/mesenchymal phenotypes upon forced expression of the indicated TFs. Successfully infected population (GFP-positive) can be distinguished from the uninfected population (GFP-negative) by GFP fluorescence (left panels). Ecad/Vim profiles are visualized separately for GFP-positive and GFP-negative populations. Analysis was performed on MCF10A cells five days after infection. Note that GFP-negative population serves as an internal control. Only GFP-positive populations were analyzed in the experiments for main figures.

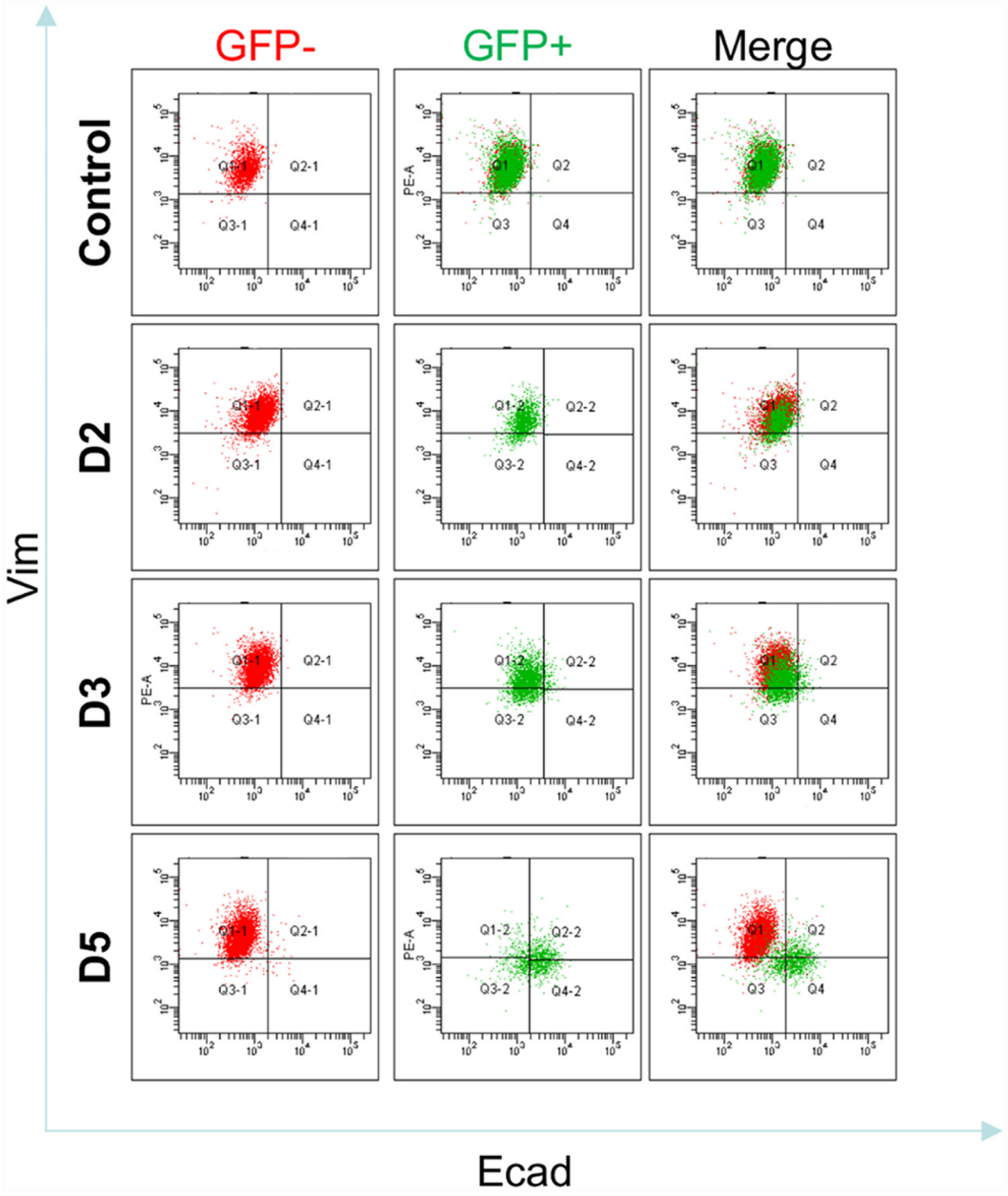

Figure H.4: Time series of Ecad/Vim profile change upon Ovol2 expression in MDA-MB231 cells. Cells were infected with Ovol2-expressing lentivirus and Ecad/Vim profile was analyzed by flow cytometry at the indicated time points. Empty vector control at day 5 is shown at the top.

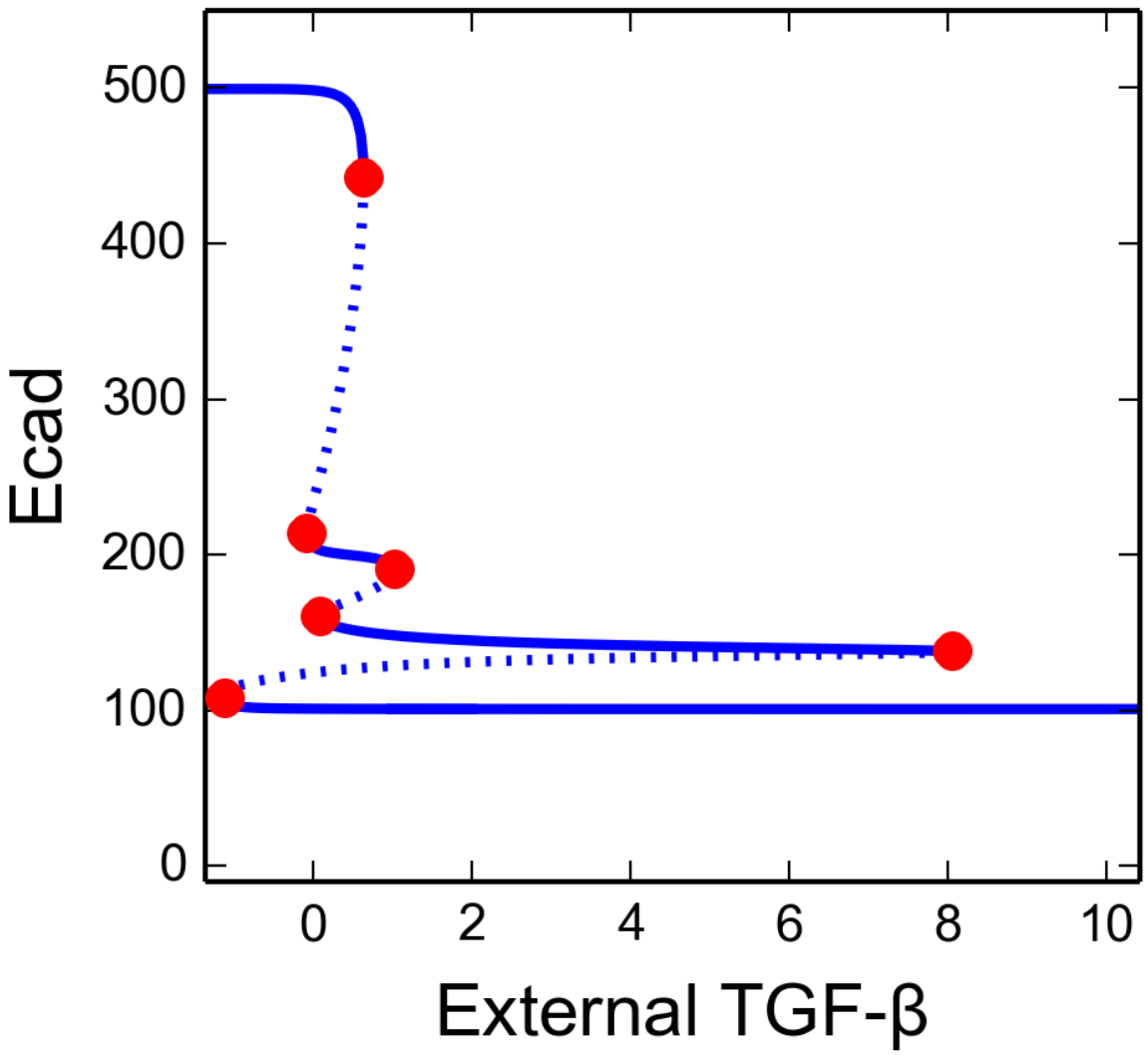

Figure H.5: Bifurcation diagram for Ecad with respect to external TGFβ. Solid curve: stable steady state. Dashed curve: unstable steady state. Red dots: saddle-node bifurcation points used for computing curves in Figure 2.5A.

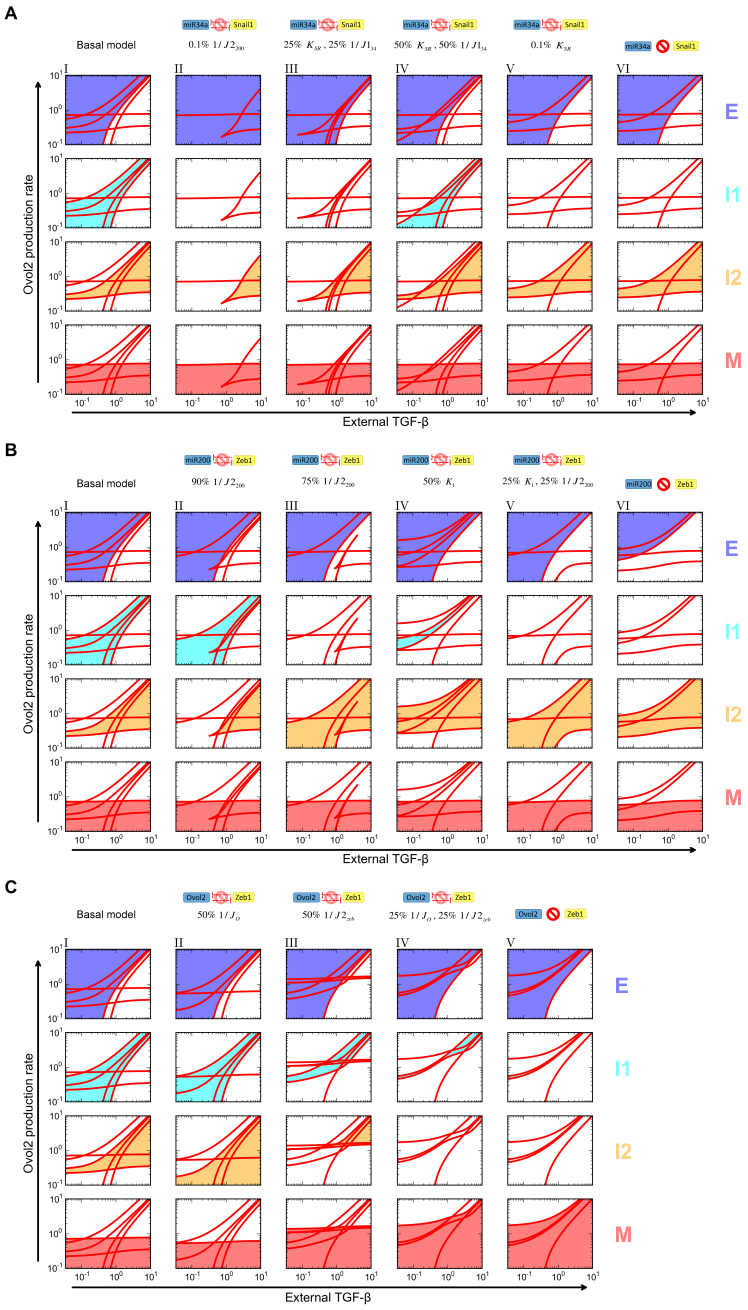

Figure H.6: Effects of partial blockage of mutual inhibition loops. Shown are comparisons of the basal model (leftmost column), complete blockage (rightmost column), and partial blockage (middle columns) of miR34a-Snail (A), miR200-Zeb1 (B) and Ovol2-Zeb1 (C) mutual inhibition loops on the four phenotypes. Each subplot is a two-parameter bifurcation diagram similar to Figure 2.5A. Subplots in each column highlight the various phenotypes in one condition. Shaded areas are highlighted phenotypes. Colors of the shading correspond to the colored labels on the right.

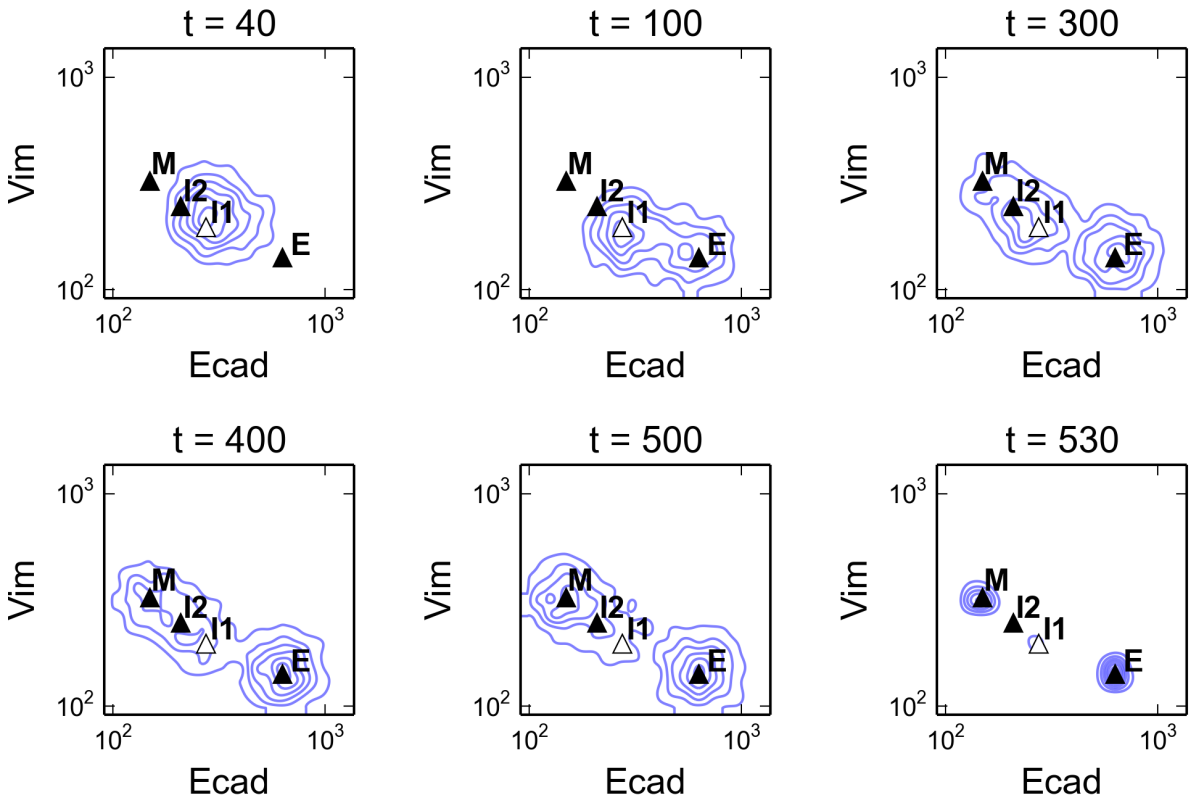

Figure H.7: Stochastic transitions from I_1 state to E and M states. Stochastic simulations for a population of 2000 cells. The basal parameter set and initial condition at I_1 were used (as in Figure 2.8). TGFβ concentration was raised from 0 to 0.5 at t = 100. White noise terms were set to zero at t = 500.