Stochastic Dynamics and Epigenetic Regulation of Gene Expression: from Stimulus Response to Evolutionary Adaptation

by

Mariana Gomez-Schiavon

Graduate Program in Computational Biology and Bioinformatics
Duke University

Date: __________________________

Approved:

Nicolas E. Buchler, Supervisor

L. Ryan Baugh

Katharina V. Koelle

Joshua E. Socolar

Dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Graduate Program in Computational Biology and Bioinformatics in the Graduate School of Duke University

2016
Abstract

Stochastic Dynamics and Epigenetic Regulation of Gene Expression: from Stimulus Response to Evolutionary Adaptation

by

Mariana Gomez-Schiavon

Graduate Program in Computational Biology and Bioinformatics
Duke University

Date: __________________________
Approved:

Nicolas E. Buchler, Supervisor

L. Ryan Baugh

Katharina V. Koelle

Joshua E. Socolar

An abstract of a dissertation submitted in partial fulfillment of the requirements for Doctor of Philosophy in the Graduate Program in Computational Biology and Bioinformatics in the Graduate School of Duke University
2016
Abstract

How organisms adapt and survive in continuously fluctuating environments is a central question of evolutionary biology. Additionally, organisms have to deal with the inherent stochasticity in all cellular processes. The purpose of this thesis is to gain insights into how organisms can use epigenetics and the stochasticity of gene expression to deal with a fluctuating environment. To accomplish this, two cases at different temporal and structural scales were explored: (1) the early transcriptional response to an environmental stimulus in single cells, and (2) the evolutionary dynamics of a population adapting to a recurring fluctuating environment. Mathematical models of stochastic gene expression, population dynamics, and evolution were developed to explore these systems.

First, the information available in sparse single cell measurements was analyzed to better characterize the intrinsic stochasticity of gene expression regulation. A mathematical and statistical model was developed to characterize the kinetics of a single cell, single gene behavior in response to a single environmental stimulus. Bayesian inference approach was used to deduce the contribution of multiple gene promoter states on the experimentally measured cell-to-cell variability. The developed algorithm robustly estimated the kinetic parameters describing the early gene expression dynamics in response a stimulus in single neurons, even when the experimental samples were small and sparse. Additionally, this algorithm allowed testing and comparing different biological hypotheses, and can potentially be applied to a
variety of systems.

Second, the evolutionary adaptation dynamics of epigenetic switches in a recurrent fluctuating environment were studied by observing the evolution of gene regulatory circuit in a population under multiple environmental cycles. The evolutionary advantage of using epigenetics to exploit the natural noise in gene expression was tested by competing this strategy against the classical genetic adaptation through mutations in a variety of evolutionary conditions. A trade-off between minimizing the adaptation time after each environmental transition and increasing the robustness of the phenotype during the constant environment between transitions was observed. Surviving lineages evolved bistable, epigenetic switching to adapt quickly in fast fluctuating environments, whereas genetic adaptation with high robustness was favored in slowly fluctuating environments.
Contents

Abstract iv
List of Tables xi
List of Figures xii
List of Abbreviations and Symbols xxxiii
Acknowledgements xxxvi

1 Introduction 1
  1.1 Overview . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 1
  1.2 Gene expression regulation . . . . . . . . . . . . . . . . . . . . . . . 3
    1.2.1 Deterministic dynamics of gene expression . . . . . . . . . . 7
    1.2.2 Stochastic dynamics of gene expression . . . . . . . . . . . 8
  1.3 Biochemical noise . . . . . . . . . . . . . . . . . . . . . . . . . . . 11
  1.4 Gene circuits and emergent properties . . . . . . . . . . . . . . . . 12
    1.4.1 Bistability . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 14
    1.4.2 Bistable switches . . . . . . . . . . . . . . . . . . . . . . . . . 14
  1.5 Genetic vs epigenetic inheritance . . . . . . . . . . . . . . . . . . . 17
  1.6 Fitness and adaptation in a static environment . . . . . . . . . . . 20
  1.7 Adaptation in fluctuating environments . . . . . . . . . . . . . . . 21
    1.7.1 Sensing and responding to a stimulus . . . . . . . . . . . . . 22
    1.7.2 Adapting through genetic mutations . . . . . . . . . . . . . 22
1.7.3 Anticipating the changes: Bet-hedging .......... 23
1.8 Simulating evolution .......................... 24
1.9 Organization of Dissertation ............... 25

2 Stochastic dynamics of early transcriptional response: Activity-dependent immediate-early gene expression in single neurons 27
2.1 Framework of stochastic dynamics of early transcriptional response .. 30
2.2 Activity-dependent, immediate-early transcriptional response .... 34
   2.2.1 Transcriptional mechanisms of activity-dependent synapse development and plasticity ..................... 34
   2.2.2 Molecular mechanisms regulating activity-dependent transcription of immediate-early genes .................. 35
   2.2.3 Single-molecule activity-dependent gene expression temporal dynamics in neurons ............................ 37
2.3 Model of stochastic dynamics of early transcriptional response .... 40
   2.3.1 Gene-expression probability distribution and its dynamics .. 40
   2.3.2 Estimating kinetic parameters: A Bayesian approach .... 45
   2.3.3 Comparing models: Information Criterion metrics ......... 50
2.4 Characterizing the kinetic parameters of Fos transcriptional response in neurons ........................................ 52
   2.4.1 2-states model with sensitive activation rate ............. 53
   2.4.2 Other models: changing the sensitive parameters ........ 54
2.5 Using synthetic data to validate estimation and propose a better sampling approach ................................ 57
2.6 Computational cost & model complexity ....................... 63
2.7 Discussion & Conclusions .......................... 66
2.8 Contributions ...................................... 69

3 Modeling evolutionary dynamics of epigenetic switches in fluctuating environments 70
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1</td>
<td>Strategies to cope with fluctuating environments</td>
<td>73</td>
</tr>
<tr>
<td>3.1.1</td>
<td>Genetic adaptation</td>
<td>73</td>
</tr>
<tr>
<td>3.1.2</td>
<td>Stochastic epigenetic switching</td>
<td>74</td>
</tr>
<tr>
<td>3.1.3</td>
<td>Inducible switching</td>
<td>74</td>
</tr>
<tr>
<td>3.1.4</td>
<td>Ignoring the environment</td>
<td>75</td>
</tr>
<tr>
<td>3.2</td>
<td>Review of previous work on the selective advantage of epigenetic switches</td>
<td>75</td>
</tr>
<tr>
<td>3.2.1</td>
<td>Assuming a static, non-evolving population</td>
<td>75</td>
</tr>
<tr>
<td>3.2.2</td>
<td>A mechanistic model</td>
<td>78</td>
</tr>
<tr>
<td>3.2.3</td>
<td>Experimental approaches</td>
<td>80</td>
</tr>
<tr>
<td>3.3</td>
<td>Our approach – A computational experiment</td>
<td>81</td>
</tr>
<tr>
<td>3.3.1</td>
<td>Mechanistic model</td>
<td>83</td>
</tr>
<tr>
<td>3.3.2</td>
<td>Evolutionary model</td>
<td>92</td>
</tr>
<tr>
<td>3.3.3</td>
<td>Systematic test of evolutionary conditions</td>
<td>96</td>
</tr>
<tr>
<td>3.4</td>
<td>Alternative assumptions</td>
<td>96</td>
</tr>
<tr>
<td>3.4.1</td>
<td>Moran model</td>
<td>97</td>
</tr>
<tr>
<td>3.4.2</td>
<td>Environmental random fluctuations</td>
<td>98</td>
</tr>
<tr>
<td>3.4.3</td>
<td>Phenotype definition</td>
<td>98</td>
</tr>
<tr>
<td>3.4.4</td>
<td>Fitness functions</td>
<td>98</td>
</tr>
<tr>
<td>3.4.5</td>
<td>Selection schemes</td>
<td>99</td>
</tr>
<tr>
<td>3.4.6</td>
<td>Mutation scheme</td>
<td>99</td>
</tr>
<tr>
<td>3.4.7</td>
<td>Optimal phenotypes</td>
<td>100</td>
</tr>
<tr>
<td>3.4.8</td>
<td>Basal activity</td>
<td>100</td>
</tr>
<tr>
<td>3.4.9</td>
<td>Degradation rate</td>
<td>101</td>
</tr>
<tr>
<td>3.5</td>
<td>Discussion &amp; Conclusions</td>
<td>101</td>
</tr>
</tbody>
</table>
4 Populations adapt to environmental fluctuations 103

4.1 Adaptation to fluctuating environments ......................... 104
  4.1.1 The “no-response” strategy is not a stable solution .......... 106
4.2 Population dynamics during adaptation ............................ 110
  4.2.1 Monostable and bistable subpopulations ..................... 116
  4.2.2 Mutational cloud effect .................................... 116
4.3 Discussion & Conclusions ........................................ 121

5 Characterizing adaptation strategies & their selective advantage 122

5.1 Individual lineages & adaptation strategies ........................ 123
  5.1.1 Tracing individual’s history .................................. 123
  5.1.2 Characterizing adaptation strategies ......................... 125
  5.1.3 Evolutionary conditions & the selected adaptation strategy .. 127
5.2 Fitness costs: Adaptation time vs Phenotypic robustness ........ 136
  5.2.1 Effect of Hill coefficient .................................... 141
5.3 Discussion & Conclusions ........................................ 143
5.4 Contributions .................................................... 147

6 Future directions 148

6.1 Stochastic dynamics of early transcriptional response .......... 148
  6.1.1 A more complex model of gene expression .................... 148
  6.1.2 A more efficient algorithm for Bayesian inference ........ 149
  6.1.3 Using prior biological knowledge to constrain Bayesian inference152
6.2 Theoretical framework for the evolution of epigenetic switches in fluctuating environments ............................. 152
6.3 Other adaptation strategies for fluctuating environments .......... 153
6.4 Considerations for the study of epigenetic switches in experimental systems ........................................ 154
A Stochastic dynamics of early transcriptional response – PIPELINE

A.1 Data ................................................................. 156
A.2 Define mathematical model ........................................ 157
A.3 Calculate probability distributions ................................. 159
A.4 Compare model & data .............................................. 160
A.5 Metropolis Random Walk algorithm .............................. 161
A.6 Synthetic data ....................................................... 167

B Previous studies on stochastic dynamics of 2-states models of gene expression 170

C Previous studies on evolution of epigenetic switches 173

D Evolutionary dynamics of epigenetic switches in fluctuating environments – PIPELINE 185

D.1 Evolutionary dynamics simulations ................................. 185
  D.1.1 Running instructions ............................................ 186
  D.1.2 EvoDynamics.cpp .............................................. 190
  D.1.3 GeneNetwork.h ............................................... 195
  D.1.4 ran1.h ...................................................... 203

Bibliography 204

Biography 215
List of Tables

3.1 Evolutionary parameters ........................................ 97
List of Figures

1.1 Central dogma of molecular biology & transcriptional regulation. (A) The central dogma of molecular biology states that the genetic information is encoded in the DNA sequence, which can copy itself – replication – or be transferred to RNA – transcription –, and then transferred to protein – translation. (B) A diagram of transcription regulation in bacteria. A constitutive gene is expressed without active regulation; the gene can also be regulated by the binding and unbinding of transcription factors, which can work as repressors blocking the binding of the RNA polymerase (RNA Pol), or as activators facilitating the binding of the RNA Pol, which transcribes the gene into messenger RNA (mRNA). (C) Analogously, transcription in eukaryote cells is regulated by transcription factors, but more components are involved, including the holoenzyme complex (orange/brown), enhancer elements, and chromatin structure regulators (not shown). Additionally, the gene sequence might contain intron sequences, which are removed after transcription, and only the exon sequences are translated.

1.2 Regulation of transcription rates. Effect of the cooperativity ($n_H$) over the transcription rate function of the concentration of (A) an activator or (B) a repressor of the gene. The constant $K_D$ is the concentration at which the transcription rate is half of the maximum rate $k$, and the Hill coefficient $n_H$ quantifies the cooperativity effect of the transcriptional factors (Equation 1.6).
1.3 **Steady state of gene expression and its stability.** A steady state \( (X^*) \) occurs whenever the synthesis and degradation rate are equal, i.e. \( dX^*/dt = 0 \). (A) An example with a stable steady state: if a perturbation around the steady state decreasing the gene expression \( X^* - \epsilon \) results in \( d(X^* - \epsilon)/dt > 0 \), and a perturbation increasing the gene expression level \( X^* + \epsilon \) results in \( d(X^* + \epsilon)/dt < 0 \); i.e., the steady state works as an attractor. (B) An example showing both a stable and an unstable steady state. The unstable steady state works as a repeller, i.e. a perturbation around this steady state is not attracted back.

1.4 **Stochastic biochemical events and stationary distribution.**
(A) The synthesis and degradation reactions are probabilistic events which change the expression level \( X \) by \( \pm 1 \) molecules. The probability of these events depends on the reaction propensities \( (f_+(X) \text{ and } f_-(X)) \) and the current expression state \( (X) \). (B) The gene expression dynamics behave as a random walk on a discrete lattice. (C) The probability distribution of gene expression reaches a stationary state when \( \partial P^*(X, t)/\partial t = 0 \forall X \).

1.5 **Gene circuit motifs.** Examples of regulatory motifs, where \( X, Y, \) and \( Z \) represent transcription factors (TFs). The green arrow from a TF to another indicates gene expression activation, while the blunt red arrow indicates repression.

1.6 **Bistability in gene expression.** (A) Diagram of a self-activating gene. The protein produced by the gene acts as a transcription factor activating its own synthesis; this is an example of a positive feedback loop. The protein is also constantly degraded. (B) Diagram of the synthesis (green thick line) and degradation (gray dashed line) rates of the transcription factor. A steady state occurs whenever the synthesis and degradation rate are equal, and the steady state is stable if it works as an attractor. Alternative synthesis rate functions (green dotted and dotdash lines) show that high nonlinearity and a specific range of regulatory parameters are necessary for the system to be bistable.
1.7 Bistability and the stationary distribution. Three examples of a self-regulated gene $X$ are shown, where only the level of nonlinearity $n_H$ of the synthesis rate varies between them (see Eq. 1.6). The synthesis ($f_+(X)$) and degradation ($f_-(X)$) propensity functions, the resulting stationary distribution ($P^*(X,t)$) and the potential energy function ($V(X) = -\int_0^{X} (f_+(x) - f_+(x))\,dx$) are shown for each case. Notice that the stable steady states roughly match the modes in $P^*(X,t)$, and these appear as basins in the $V(X)$, while the unstable steady states are hills. At certain point, increasing the nonlinearity of the synthesis rate does not affect the deterministic steady states but can still have a big effect on the stationary distribution.

1.8 Stochastic transitions in epigenetic switches. (A) Potential energy function of a bistable system with stable steady states $B_L$ and $B_H$, and unstable steady state $B_U$. (B) The gene expression dynamics of the same bistable system including biochemical noise. Spontaneous transitions occur between “stable” steady states. In the stationary state, the gene expression distribution is a bimodal distribution with the modes matching $B_L$ and $B_H$.

1.9 Induced transitions in epigenetic switches. (A) Diagram of a bistable (epigenetic) switch dependent of a stimulus. One or two stable states exist in the system as a function of the stimulus level. In the bistable region two stable (thick line) and one unstable (dashed line) coexist. Overlooking the biochemical noise, the system is induced (green dotted line) only if the stimulus level is higher than $s_I$; but if the stimulus level is decreasing, the system “remembers” the previous state and only transitions to the uninduced state (red dotted line) if the stimulus is less than $s_U < s_I$. This shift between the induction and uninduction threshold ($s_I \neq s_U$) is known as hysteresis. (B) Diagram of the expression of the transcriptional factor of the epigenetic switch as the stimulus level changes over time, overlooking the biochemical noise. The potential energy function corresponding to different levels of stimulus, as well as the state of the system in these (blue dot), are shown as a reference. Notice that the system is bistable always that the stimulus $s$ is $s_U \leq s \leq s_I$, but the displayed state depends on the cellular history.
2.1 **Diagram of 2-states model of mRNA expression.** Each cell has two copies of the gene, and each gene promoter (circles) can be either in active ($\rho_{ON}$) or inactive ($\rho_{OFF}$) state; an active promoter turns inactive with rate $k_{OFF}$, and an inactive promoter turns active with rate $k_{ON}$. Each promoter copy synthesizes mRNA molecules ($m$) with rate $\mu$ or $\mu_0$ if it is active or inactive, respectively; and the mRNA molecules are degraded with rate $\delta$. The stimulus ($s$) can potentially affect each of these molecular processes.

2.2 **Effect of parameters on gene expression and bursts dynamics in single cell.** Diagram exemplifying the aspects of the dynamical system affected by each individual parameter in the model: $k_{ON}$ determines the frequency of burst, i.e. the expected time the promoter stays in the inactive state (e.g. blue line); $k_{OFF}$ determines the length of the bursts, i.e. the expected time the promoter stays in the active state (e.g. violet line); $\mu$ determines the magnitude of the burst, i.e. how much mRNA is accumulated once a promoter is in the active state (orange arrow); $\mu_0$ determines a basal mRNA signal (brown bar); and $\delta$ defines how the burst decays over time (gray arrow). The parameter values used are: $k_{ON} = 0.01 \text{ min}^{-1}, k_{OFF} = 0.02 \text{ min}^{-1}, \mu_0 = 0.1 \text{ mRNA/min}, \mu = 5 \text{ mRNA/min}$, and $\delta = 0.0462 \text{ min}^{-1}$. The graph shows the simulated dynamics of one cell, where $\rho_1, \rho_2$ are the individual promoter copies and the color reflects their state (green if active, $\rho_{ON}$; gray if inactive, $\rho_{OFF}$), and mRNA is the number of $m$ molecules over time. The average mRNA number in the particular simulation is also shown as a dashed yellow line.

2.3 **Increasing gene expression average level.** Examples of how the system dynamics are affected as individual parameters change. Same graphs and notation as Figure 2.2; the same parameters were used except for: (A) higher $k_{ON} \rightarrow 0.04 \text{ min}^{-1}$, (B) lower $k_{OFF} \rightarrow 0.005 \text{ min}^{-1}$, (C) higher $\mu \rightarrow 10 \text{ mRNA/min}$, and (D) higher $\mu_0 \rightarrow 1 \text{ mRNA/min}$. In all cases, the average mRNA levels increased with respect to Figure 2.2, but very different dynamics are observed.
2.4 Population distributions display signatures of the gene expression dynamics. (A) A temporal snapshot (red line) of the population. (B) The stationary probability distribution of the population shown in (A). The corresponding biophysical parameters are shown in the bottom, and the expected average mRNA molecules number in the population $\langle x_{\text{mRNA}} \rangle$ is 75. (C) Examples of how the stationary probability distributions change as individual parameters vary: $k_{\text{ON}} \rightarrow 0.04 \text{min}^{-1}$, $k_{\text{OFF}} \rightarrow 0.005 \text{min}^{-1}$, $\mu \rightarrow 10 \text{mRNA/min}$, and $\mu_0 \rightarrow 2.5 \text{mRNA/min}$. Notice that the expected average mRNA molecules number in the population is practically identical in all these cases $\langle x_{\text{mRNA}} \rangle \approx 150$, while the populations show qualitatively different behavior (see Figure 2.2).

2.5 Mechanisms regulating activity-dependent early gene transcription in neurons. A simple diagram of the diverse mechanisms involved in the regulation of activity-dependent immediate-early gene transcription. Before neural activity, the promoter is primed for transcription with transcription factors (TF1,2) pre-bound to stimulus-response elements (gray boxes) and distal enhancer elements (EE), and the RNA-polymerase II (Pol II) paused in the promoter; additionally, some enzymes (E1) help to keep the repressed promoter state (e.g. histone deacetylase). Some leaky transcription (thin black arrows) might occur –which is expected in both directions–, but the promoter is considered inactive (blunt red arrow). After neural activity –increase in intracellular calcium ([Ca$^{2+}$])–, some enzymes are released (E1), and others recruited (E2), making posttranslational modifications to both transcription factors (e.g. phosphorylation, P) and histones (e.g. acetylation, A), which induce gene transcription (thick black arrow). Long-distance looping of DNA and the proximity of associated enhancers (EE) have also been proposed as a mechanism of regulation of this activity-dependent transcription; in particular, eRNAs –short non-coding RNAs transcribed from enhancer regions– have been detected under neural activity. Now, the promoter is considered active (green arrow).
2.6 **Single-molecule mRNA fluorescence in situ method (smFISH).** Diagram of the smFISH method. Fluorescent labeled probes bind specifically to the gene of interest, either in the intron regions –detecting only nascent mRNAs– or in the exon regions –detecting both nascent and mature mRNAs–. The nascent mRNAs accumulate in active promoters, appearing as intense dots in the nucleus –one for each active promoter–; if two-colored labeled probes are used –one for the introns, other for exons– the resulting intense dots are expected to co-localize. Individual mature mRNA molecules appear as small fluorescent dots in the cytoplasm.  

2.7 **Single-molecule fluorescent in situ hybridization (smFISH) measurements of Fos mRNA in neurons.** Measurements are shown before the stimulus (uninduced), after 5 minutes in KCl stimulus, and additional 10 minutes and 20 minutes in common condition medium. (A) An example of an image-processed cell for each condition is shown, highlighting the detected mRNAs (green circles) and active transcription sites (TSs; green squares). The contour of the considered cell is marked (yellow line), as well as its nucleus (dashed yellow line). (B) For each condition, the histogram of number of free mRNA molecules detected per cell with a particular number of active transcription sites (TSs) are shown (blue dots, left y-axis), as well as a smoothed histogram with bins ±10 mRNA number (gray dots, right y-axis). The total number of cells per time sample is listed at the top (sample size, n).  

2.8 **Possible reactions and system state space.** The cell's state $x$ change over time is equivalent to a random walk on the lattice depicted above. The individual cell state $x$ is defined as the number of active promoters $\rho_{ON} \in \{0, 1, 2\}$ and the number of mRNA molecules $m \in \{0, 1, 2, \ldots\}$. There are four possible reactions: (1) Promoter activation (blue arrow), which increases $\rho_{ON}$ by one; (2) promoter inactivation (purple arrow), which decreases $\rho_{ON}$ by one; (3) mRNA synthesis (orange arrow), which increases $m$ by one; and (3) mRNA degradation (gray arrow), which decreases $m$ by one.
2.9 Example of a specific model for the activity-dependent gene expression system. An example with only $k_{ON}$ sensitive to the stimulus, and biophysical parameter values: $k_{OFF} = 0.0459$ min$^{-1}$, $\mu_0 = 0.026$ mRNA/min, $\mu = 5.65$ mRNA/min, $k_{ON}^{(U)} = 0.0049$ min$^{-1}$, $k_{ON}^{(S)} = 0.1110$ min$^{-1}$. For the uninduced condition and each time point measurement after stimulus the probability distributions (pink lines), and an example of a random sample $Y$ of 150 cells per condition (blue dots), are shown.  

2.10 Example of a run of the Metropolis Random Walk algorithm in our system. (A) The log-likelihood per iteration, highlighting the “burn-in” period (red dotted line). (B) The selected parameter values per iteration after the “burn-in” period. In this example, the 2-states $k_{ON}$-sensitive model is used (see Appendix A for implementation details).  

2.11 Biophysical parameter inference under a 2-states $k_{ON}$-sensitive model. (A) Marginal posterior distributions for the biophysical parameter values for three MRW replicas; $k_{ON}^{(U)}$ refers to the $k_{ON}$ value in uninduced conditions ($\theta_U$) and $k_{ON}^{(S)}$ refers to the $k_{ON}$ value in after stimulus conditions ($\theta_S$). (B) Joint distributions of pairs of parameter values. (C) Joint distribution of parameter values and log-likelihood. (B-c) use the results from all replicas.  

2.12 Variance in parameter inference under a 2-states $k_{ON}$-sensitive model. Fold change variation and coefficient of variation (i.e. variance over mean) observed in the individual parameters in the posterior distributions from Figure 2.11. In each boxplot, the central mark is the median, the edges of the box are the 25th and 75th percentiles, the whiskers extend to the most extreme data points not considered outliers, and outliers are plotted individually.  

2.13 Probability distribution associated with parameter estimates under a 2-states $k_{ON}$-sensitive model. Resulting variation in the corresponding gene expression probability distributions for the $\theta$’s obtained in the MRW from Figure 2.11. The experimental data $Y = \{y_i\}$ is shown for comparison (gray dots).  

2.14 Comparing parameter distributions of different 2-states models. The marginal posterior distribution for each biophysical parameter. Each row corresponds to a different model with the parameter(s) sensitive to stimulus shown in the left. When the parameter is sensitive to stimulus, uninduced conditions are shown as continuous lines, after stimulus conditions as dashed lines. The results from three MRW replicas are used in all cases.
2.15 **Comparing different 2-states models by Information Criteria.** Three distinct metrics are applied – *Bayesian Information Criterion* (BIC), *Akaike Information Criterion* (AIC), and *Deviance Information Criterion* (DIC) – to the results obtained from the MRW to 2-states model where only the parameter(s) listed in the x-axis are sensitive to the stimulus.

2.16 **Biophysical parameter inference for $k_{ON}, k_{OFF}$, $\mu$-sensitive model.** Marginal and joint posterior distributions for the log-likelihood and biophysical parameter values for three MRW replicas; $p^{(U)}$ refers to the parameter value in *uninduced* conditions ($\theta_U$) and $p^{(S)}$ refers to the parameter value in *after stimulus* conditions ($\theta_S$) in each case.

2.17 **Biophysical parameter inference for the $k_{ON}, k_{OFF}$-sensitive model.** Marginal and joint posterior distributions for the log-likelihood and biophysical parameter values for three MRW replicas; $k_{OFF}^{(U)}$ refers to the $k_{OFF}$ value in *uninduced* conditions ($\theta_U$) and $k_{OFF}^{(S)}$ refers to the $k_{OFF}$ value in *after stimulus* conditions ($\theta_S$).

2.18 **Sample strategy effect on posterior probability distributions.** The marginal posterior probabilities per biophysical parameter per sample strategy are shown for three random samples. The results from three independent MRW runs – different random initial conditions – were mixed for each case, excluding their “burn-in” period. In all cases, the x- and y-axes have exactly the same limits in each column and row, respectively, and the true parameter value $\hat{\theta} \in \theta$ are shown as a dashed vertical line.

2.19 **Mean deviation $\langle \epsilon \rangle$ per sample strategy.** The mean deviation of the posterior distribution $\langle \epsilon \rangle$ per sample strategy (Eq. 2.20) for three random samples (gray circles) are shown; the results from three independent MRW runs – different random initial conditions – were mixed for each case, excluding their “burn-in” period. The average of the independent samples is also shown for each strategy (squares): replicating the experimental sample (red), increasing the sample size (orange), and 600 cells in the total sample varying the time point measurements (green).
3.1 Genetic adaptation versus epigenetic switching of a self-activating gene. (A) Diagram of a self-activating gene that considers two biochemical events: protein synthesis with a rate that increases with number of proteins \( A \) (i.e., positive feedback loop) and protein degradation. On the right, a cartoon of the biophysical parameter space or genotypes (\( \theta \)) of this gene circuit with two characteristic regions: monostable (white) and bistable (pink) phenotypes. In the monostable region, a genotype might be optimal in either one environment (e.g. LOW protein numbers, \( \theta_L \)) or the other (e.g. HIGH protein numbers, \( \theta_H \)). Genetic mutations are required to change from one solution the other (blue arrow). In the bistable region, a single genotype (e.g. \( \theta_B \)) can display two different phenotypes with each phenotype potentially optimal in both environments. (B) Cartoon of the protein number \( A \) dynamics in an individual cell with each of the genotypes described in (A). Monostable genotypes (\( \theta_L \) and \( \theta_H \)) exhibit a unimodal distribution of protein expression (\( \rho(A) \)), whereas a bistable genotype (\( \theta_B \)) exhibits a bimodal distribution having spontaneous transitions between phenotypic states over time (i.e. epimutations) triggered by stochastic gene expression. Cartoon of the population dynamics using (C) genetic adaptation or (D) epigenetic switching to adapt after an environmental change. The fitness score function (\( \omega \); orange dashed line) and the phenotype distribution of the population (\( P(A) \); blue line) are shown for each generation (\( g \)), and the fraction of the population expected to be selected in the next generation (i.e. individuals with higher fitness scores) are highlighted (blue area). The environment changes from selecting HIGH protein numbers (light green) to select LOW protein numbers (dark green).

3.2 Monostable and bistable solutions. Different parametrizations of the model are shown, one resulting in a monostable solution (left), and other one displaying bistability (right). The synthesis (green line) and degradation (gray dashed line) rate functions are shown; whenever these two functions intersect, a steady state occurs. The corresponding potential energy function \( V(A) = - \int_0^a (f(a) - \gamma \cdot a) \, da \) is shown in the bottom. The stable steady states appear as basins in \( V(A) \), and work as attractors (i.e. after any small perturbation, the system is attracted back to the steady state); while the unstable steady state in the bistable example corresponds to a hill in the \( V(A) \) function, and any small perturbation will pull the system to one of the attractors. As a reference, the arrows show the direction of the change on \( A \) (i.e. \( \partial A/\partial t \) sign) around the steady states.
3.3 **A monostable system can display bimodality.** An example showing a bimodal protein number stationary distribution $P^*(A)$ (the local maxima are shown as black circles), even if only one steady state solution (i.e. $\partial A/\partial t = 0$) exists.

3.4 **Biophysical parameters, deterministic steady state solutions and stochastic stationary distributions of protein levels.** (A) The effect of the maximum synthesis rate ($k$) and the affinity constant ($K_D$) over the deterministic steady state solutions of the protein expression (i.e. $\frac{dA^*}{dx} = f(A^*) - A^* = 0 \iff f(A^*) = A^*$ where $f(A) = k(\alpha + (1 - \alpha)\frac{A^{n_H}}{A^{n_H} + K_D^{n_H}})$) in the limit of high Hill coefficients ($n_H \to \infty$). If $K_D < \alpha k$ the system is monostable HIGH with the protein expression steady state ($A^*$) equal to $k$; on the other hand, if $K_D > k$ then the system is monostable LOW with $A^* = \alpha k$. When $\alpha k \leq K_D \leq k$ is intermediate, these two steady states coexist and the system is bistable. (B) Bifurcation diagram of the protein steady states as the Hill coefficient ($n_H$) varies while keeping the rest of the biophysical parameters fixed. As $n_H$ value increases, the system goes from monostable (blue dots) to bistable (violet and pink dots). As $n_H \to \infty$, the stable steady states monotonically approach their limiting values, $\alpha k$ and $k$ (dashed gray lines), and the unstable steady state asymptotically approaches $K_D$ (dotted gray line). I show a few examples of the stationary distribution of the protein expression for stochastic simulations with intrinsic biochemical noise (bottom). As $n_H$ approaches the bifurcation point (where the system passes from being monostable to bistable) the stationary distribution becomes wider (i.e. the phenotype is more variable). In the bistable region, even if the two modes of the stationary distribution do not change much, their relative weights can be significantly affected by the value of the unstable steady state, as stochastic transitions from one stable mode to the other become more or less probable.
3.5 **Evolutionary model.** The environment fluctuates periodically with frequency $\nu$. The total number of generations spent in a constant environment (epoch) has the same length $(1/\nu)$ and each environment (HIGH or LOW) selects for a different distribution of protein levels (phenotypes). Each generation, I simulated the stochastic protein dynamics of a self-activating gene in each cell across a population of size $N$. At the end of each simulation, the population phenotypes varied because gene expression is stochastic and because cells can have different underlying biochemical parameters. The current environment in each generation assigned a fitness ($\omega$) to each cell based on its final protein level. I used tournament selection (where $s_t$ determines the strength of selection) to determine the next generation of cells according to their fitness. Each cell in the next generation was mutated with probability $u$, where the current set of biophysical parameters were multiplied or divided up to a maximum step-size of $M$. . . . . . . . . 93

4.1 **Fitness as a population adapts to a fluctuating environment.**
The initial population started from a non-optimal genotype ($\theta_0$) where $k = 80$, $n_H = 1$, $K_D = 10$ with evolutionary parameters $N = 10000$, $\nu = 0.1$, $s_t = 6$, $u = 0.03$, and $M = 1.1$. The geometric mean of the population fitness per environmental cycle ($W_{cycle}$) is shown, where each cycle spans a LOW (dark green) and HIGH (light green) epoch. The initial genotype was well adapted only to the HIGH environment, the population eventually evolved to a solution with higher $W_{cycle}$. This final genotype had high population fitness ($w$) in both environments and rapidly adapted after each environmental transition. 105

4.2 **Average genotype in a population adapting to the fluctuating environment.** The initial population started from a non-optimal genotype ($\theta_0$) where $k = 80$, $n_H = 1$, $K_D = 10$ with evolutionary parameters $N = 10000$, $\nu = 0.1$, $s_t = 6$, $u = 0.03$, and $M = 1.1$. The average value in the population of each biophysical parameter ($\langle \theta \rangle_g = \langle k_g, n_{H,g}, K_{D,g} \rangle$) is shown for each generation. . . . . . . . . 106
4.3 **Population fitness and nonlinearity per cycle.** The initial population started from a non-optimal genotype ($\theta_0$) where $k = 80$, $n_H = 1$, $K_D = 10$ with evolutionary parameters $N = 10000$, $\nu = 0.1$, $s_t = 6$, $u = 0.03$, and $M = 1.1$. The geometric mean of the population fitness per environmental cycle ($W_{\text{cycle}}$) versus the average Hill coefficient per cycle ($\langle n_H \rangle_{\text{cycle}}$). Each cycle spans a LOW and HIGH epoch and there are 500 environmental cycles (increasing from gray to black) over 10,000 generations for this simulation. The insets show the population fitness per generation for the highlighted cycles (first, 89th, and last cycles), with the color bar in the top specifying the environmental state (LOW, dark green; HIGH, light green). The plot in the bottom is a close-up of the early cycles.

4.4 **Counteracting selection pressure destabilizes the “no-response” genotype in a fluctuating environment.** (A) A population with an initial synthesis rate $k = 80$, and fixed Hill coefficient $n_H = 1$, and constant $K_D = 10$, was evolved with parameters $N = 100$, $\nu = 0.1$, $s_t = 6$, $u = 0.03$, and $M = 1.1$. The population average per generation $\langle k \rangle_g$ and population fitness $w$ (dark green for LOW, light green for HIGH) are shown; geometric mean fitness per cycle ($W_{\text{cycle}}$) is also shown (orange). A sample evolutionary trajectory over one cycle is shown in red, starting from the white-filled circle. (B) A single Metropolis walker with no biochemical noise in the same fluctuating fitness landscape, where a mutation ($\theta'$) occurs every generation ($u = 1$) and it is accepted (i.e. $\theta \leftarrow \theta'$) with probability $\min(1, w^{(F)}(\theta')/w^{(F)}(\theta))$. The maximum synthesis rate per generation $\langle k \rangle_g$ and individual’s fitness $w$ (dark green for LOW, light green for HIGH) are shown; geometric mean fitness per cycle ($W_{\text{cycle}}$) is also shown (orange). A sample evolutionary trajectory over one cycle is shown in red, starting from the white-filled circle.

4.5 **Fitness advantage of increasing $K_D$ in low nonlinearity regimes.** (A) Same approach and plot as in Figure 4.4A, but with a $K_D = 20$. (B) Comparing $W_{\text{cycle}}$ distribution for $N = 100$ populations with $n_H = 1$ and $K_D = \{10, 15, 20\}$, we observed that $W_{\text{cycle}}$ gradually increases as $K_D$ increases; the same simulation but in the bistable region ($n_H = 6$; $K_D = 46$) shows a significantly higher $W_{\text{cycle}}$. . . . . . . . 110
4.6 Population dynamics in different periods of the adaptation to the fluctuating environments. Population dynamics for initial, intermediate, and final cycles in the evolutionary dynamics ($\theta_0 = \{k = 80, n_H = 1, K_D = 10\}$) with parameters $N = 10000$, $\nu = 0.1$, $s_t = 6$, $u = 0.03$, and $M = 1.1$). At each generation, I plot the joint distribution of the actual phenotype ($A$) from stochastic simulation and the expected steady state value(s) $A^*(\theta)$ from the deterministic solution given the individual’s genotype. The optimal phenotype given the environment is shown as dotted line, and the fraction of bistable individuals in the population ($f_B$) is listed in the top. The population fitness $w$ per generation for the corresponding cycles is shown on the right. 112

4.7 Average genotype in a population adapting with higher mutation step sizes and environmental fluctuation frequencies. The $W_{\text{cycle}}$, the fraction of bistable individuals ($f_B$), and the average biophysical parameters in the population are shown for each simulation, with initial genotype ($\theta_0$) $k = 80$, $n_H = 1$, $K_D = 10$, and evolutionary parameters $N = 10000$, $s_t = 6$, and $u = 0.03$. (A) $M = 2.1$ and $\nu = 0.04$ (25-generation epoch, moderate mutation step-size). (B) $M = 5$ and $\nu = 0.01$ (100-generation epoch, large mutation step-size). The population average is shown as black dotted lines, while the average on the bistable (violet) and monostable (blue) subpopulations are shown as dots for each generation. If the subpopulation was less than 1% of the population, its average genotype is excluded from the plots. 113

4.8 Average genotype in a population adapting to the fluctuating environment in the absence of biochemical noise. The $W_{\text{cycle}}$, the fraction of bistable individuals ($f_B$), and the average biophysical parameters in the population are shown for each simulation, with initial genotype ($\theta_0$) $k = 80$, $n_H = 1$, $K_D = 10$, evolutionary parameters $N = 10000$, $s_t = 6$, and $u = 0.03$, and deterministic gene expression (i.e. CONTROL simulations). (A) $M = 2.1$ and $\nu = 0.04$. (B) $M = 5$ and $\nu = 0.01$. The population average is shown as black dotted lines, while the average on the bistable (violet) and monostable (blue) subpopulations are shown as dots for each generation. If the subpopulation was less than 1% of the population, its average genotype is excluded from the plots. 114
4.9 **Average genotypes per environmental state.** The color maps show the average value of the maximum synthesis rate $k$, the Hill coefficient $n_H$, and the constant $K_D$ in the population in each environment (HIGH and LOW) of 10 independent evolutionary simulations for distinct mutation step-sizes ($M$) and environmental fluctuation frequencies ($\nu$). Each simulation ran for 10,000 generations with evolutionary parameters $N = 10000$, $s_t = 6$, $u = 0.03$, and $k = 80$, $n_H = 6$, and $K_D = 45$ as the initial genotype ($\theta_1$).  

4.10 **Coexistence of bistable and monostable subpopulations is an evolutionarily stable state.** (A) The color maps show the average fraction of bistable individuals in the population $\langle f_B|E\rangle_{\text{sim}}$ in each environment (HIGH and LOW) of 10 independent evolutionary simulations for mutation step-size ($M$) and environmental fluctuation frequency ($\nu$). Each simulation ran for 10,000 generations with evolutionary parameters $N = 10000$, $s_t = 6$, $u = 0.03$, and $k = 80$, $n_H = 6$, and $K_D = 45$ as the initial genotype ($\theta_1$). (B) Density plot of $f_B|E$ for $\nu = 0.1$, and $M = 5$ (red box in A) as a function of generations after environmental change. Each column corresponds to the distribution of the $f_B|E$ over the entire simulation. (C) Increasing the population size to $N = 25000$ sharpened the observed trends of stable co-existence. (D,E) Two examples showing how the (D) monostable and (E) bistable subpopulations can become extinct (i.e. $f_B = 1$ and $f_B = 0$, respectively; red arrows) yet are re-established in the evolutionary simulation. Both examples were for evolutionary parameters $N = 630$, $s_t = 6$, $u = 0.03$, $\nu = 0.1$, and (D) $M = 1.7$, (E) $M = 5$. I deliberately decreased $N$ because extinction events are more common in smaller populations.  

4.11 **Coexistence of bistable and monostable subpopulations in the final adapted population under multiple evolutionary conditions.** Population structure and evolutionary dynamics of adaptation to two environmental states (LOW, dark green; HIGH, light green) at 10,000 generations. Each simulation started from the same genotype and evolutionary parameters as Figure 4.7 with (A) $\nu = 0.04$, $M = 2.1$ (25-generation epoch, moderate mutation step-size) and (B) $\nu = 0.01$, $M = 5.0$ (100-generation epoch, large mutation step-size). The final adapted population contained both monostable and bistable subpopulations with bistable fraction $f_B$ (shown in pink). For each subpopulation (bistable, top; monostable, bottom), we plot the joint distribution of the actual phenotype ($A$) and deterministic phenotype $A^*(\theta)$. As a reference, the optimal phenotypes for each generation are shown as a dotted line.
4.12 Mutational cloud of the bistable subpopulation seeds the monostable subpopulation before an environmental change. A population with an initial synthesis rate \( k = 80 \) and \( K_D = 45 \), and fixed Hill coefficient \( n_H = 6 \) was evolved with parameters \( N = 10000 \), \( \nu = 0.1 \), \( s_t = 6 \), \( u = 0.03 \), and \( M = 2 \). Every generation in a constant environment, \textit{de novo} mutants arise from the bistable subpopulation which lay in monostable region with an optimal or close to phenotype in the alternative environment, which can be selected once the environment changes. The genotypes selected in the previous generation are shown as filled circles: violet if the genotype is bistable, dark blue if monostable. The \textit{de novo} mutants arising in the current generation are shown as empty circles, and the color corresponds to the parental genotype: pink if the mutant arose from a bistable parent, blue otherwise. The current environment is shown in the top of each graph, and the bistable region is marked by gray lines as reference. An example of a transition from LOW to HIGH environments is shown in the top row, and from HIGH to LOW in the bottom.

5.1 Lineage analysis of cells evolving in a fluctuating environment. At the end of each cycle (LOW epoch + HIGH epoch), I analyzed the genealogy of cells over the past two cycles. All cells were classified based on the evolutionary strategy used by their 2-cycle ancestor over a full cycle (bigger dots). The top bars show the environmental state per epoch (dark green for LOW, light green for HIGH). I plot the number (#) of distinct lineages (solid line) and genotypes (dotted line) as a function of past generations on the top row. The middle rows plot the corresponding genotypes \( \theta \) and the bottom shows the individual ancestral lineages. Ancestral genotypes can be bistable (violet) or monostable (blue). (A) Example of lineage analysis of cells that use epigenetic switching (ES) strategy for \( \nu = 0.1 \) and \( M = 5 \), i.e. their 2-cycle ancestors were fully bistable and persisted a full cycle without mutations. Note that there are distinct lineages with identical genotypes. (B) Example of lineage analysis of cells that use bistable adaptation (BA) strategy for \( \nu = 0.04 \) and \( M = 2.1 \), i.e. their 2-cycle ancestors were fully bistable but accumulated mutations over the next cycle. (C) Example of lineage analysis of cells that use genetic adaptation (GA) strategy for \( \nu = 0.01 \) and \( M = 5 \), i.e. their 2-cycle ancestors had monostable genotypes and accumulated mutations over the next cycle. In all cases, I used \( N = 4000 \), \( s_t = 15 \), and \( u = 0.03 \).
5.2 **Example of the population distribution of genotypes versus surviving lineages using ES strategy.** The first environmental cycle from Figure 5.1A is shown. The distribution of the selected population in the genotype space with $k = 80 \pm 5$ is plotted. Scale bar is set to saturate at 10% of the population (dark blue) and one individual already appears as a yellow square. The *de novo* mutations, which appear each generation and which have not yet experienced selection, are shown as small black squares. The distribution of surviving lineages genotypes are also shown (red squares). The bistable region is delimited by a pink line. Each row corresponds to the first four generations after the transition to the LOW (dark green, top) and HIGH environment (light green, bottom), as well as the last two generations in each epoch.

5.3 **Example of the population distribution of genotypes versus surviving lineages using BA strategy.** The first environmental cycle from Figure 5.1B is shown using the same notation as Figure 5.2.

5.4 **Example of the population distribution of genotypes versus surviving lineages using GA strategy.** The first environmental cycle from Figure 5.1C is shown using the same notation as Figure 5.2.

5.5 **Adaptation strategies.** Illustration of epigenetic switching (ES), bistable adaptation (BA), and genetic adaptation (GA) strategies and underlying genotypes with fixed synthesis rate ($k = 80$). The bistable region of the genotype space is highlighted in pink. The phenotype distribution $\rho(A)$ for each genotype ($\theta$) is shown in the inset, both in linear and logarithmic scale. BA is bistable, as seen in logarithmic scale, but appears effectively monostable in linear scale. This arises because $K_D$ evolves each epoch to favor one mode over the other by decreasing the relative rates of epigenetic switching between the largest and smallest mode.
5.6 Each adaptation strategy is favored under different evolutionary conditions and the transition between selected strategies is gradual. In the left, each color map shows the fraction of parental lineages using a specific adaptation strategy (epigenetic switching, ES; bistable adaptation, BA; or genetic adaptation, GA) averaged over all cycles and ten independent replica simulations for the corresponding mutation step-size \(M\) and environmental fluctuation frequency \(\nu\). Each simulation ran for 10000 generations with evolutionary parameters \(N = 10000, s_t = 6, u = 0.03\) and \(k = 80, n_H = 6\) and \(K_D = 45\) as the initial genotype \((\theta_1)\). This initial genotype sped up evolutionary simulations by being closer to final selected genotypes in all simulations. In the middle column, the equivalent results of the CONTROL simulations, where gene expression dynamics are deterministic and no stochastic epigenetic switching can occur, are shown. All lineages exhibited GA and neither bistable strategy (ES or BA) was ever selected. The corresponding bistable fraction \((f_B)_\text{sim}\) averaged over all cycles and ten independent replica simulations for the stochastic simulations (top) and deterministic CONTROL (bottom) are shown in the right.

5.7 Different adaptation strategies co-exist and exhibit large fluctuations over evolutionary time. Each plot shows the fraction of parental lineages with different adaptation strategies (epigenetic switching, ES; bistable adaptation, BA; genetic adaptation, GA) per environmental cycle for mutation step-size \(M\) and environmental fluctuation frequency \(\nu\). In each case, only the last 50 cycles are shown. Each simulation ran for 10,000 generations with evolutionary parameters \(N = 10000, s_t = 6, u = 0.03\), and \(k = 80, n_H = 6\), and \(K_D = 45\) as the initial genotype \((\theta_1)\).

5.8 Increasing selection pressure or mutation rate favors genetic adaptation, whereas increasing population size favors epigenetic switching. Each color map shows the average fraction of parental lineages using each adaptation strategy (epigenetic switching, ES; bistable adaptation, BA; genetic adaptation, GA) for the same range of mutation step-size \(M\) and environmental fluctuation frequency \(\nu\) as Figure 5.6. Evolutionary parameters used in Figure 5.6 \((s_t = 6, N = 10000, u = 0.03)\) are highlighted in red boxes. (A) The effect of only changing the selection pressure \((s_t)\) over three evolutionary replicas. (B) The effect of only changing the population size \((N)\) over three evolutionary replicas. (C) The effect of only changing the mutation rate \((u)\) over ten evolutionary replicas. All simulations ran for 10000 generations with \(k = 80, n_H = 6, K_D = 45\) as the initial genotype \((\theta_1)\).
5.9 Transitions between adaptation strategies as a function of evolutionary parameters. (A) An example showing how the current and previous adaptation strategies were defined for a surviving population. In this particular simulation, the population displayed multiple strategies in the ancestral lineage between the 2-cycle and 1-cycle ancestors (current adaptation strategy): 63.825% epigenetic switching (ES), 35.95% bistable adaptation (BA), and 0.225% genetic adaptation (GA). However, the ancestral lineage between the 3-cycle and 2-cycle ancestors (previous adaptation strategy) were 100% ES. (B) The color maps show the percentage of ancestral lineages that displayed one adaptation strategy (current adaptation strategy) and other adaptation strategy in the preceding ancestral lineage (previous adaptation strategy). These statistics were calculated for 10 evolutionary replicas for mutation step-size ($M$) and environmental fluctuation frequency ($\nu$). Each simulation was run 10,000 generations with evolutionary parameters $N = 10000$, $s_t = 6$, $u = 0.03$, and $k = 80$, $n_H = 6$, and $K_D = 45$ as the initial genotype ($\theta_1$).

5.10 The same qualitative trends on the selection of adaptation strategies per evolutionary condition is maintained in a wide variety of alternative model assumptions. Each color map shows the population average fraction of parental lineages using each adaptation strategy (epigenetic switching, ES; bistable adaptation, BA; genetic adaptation, GA) for the same range of mutation step-size ($M$) and environmental fluctuation frequency ($\nu$) as Figure 5.6. Differences in assumptions or parameters are listed above each plot. All values are the average of 3 evolutionary replicas of simulations run 10,000 generations with $N = 4,000$, $s_t = 6$, $u = 0.03$ and $k = 80$, $n_H = 6$, and $K_D = 45$ as the initial genotype ($\theta_1$). The exceptions are the weighted and proportional selection schemes where the selection pressure ($s_t$) cannot be tuned, and the Moran model case which ran for 1,000 generations. When the basal activity ($\alpha$) was changed, I adjusted the low optimal phenotype such that the ratio of $A^{(L)} = \alpha \cdot A^{(H)}$, where $A^{(H)} = 80$. See Section 3.4 for an explicit description of assumptions and parameters.
5.11 Allowing basal activity ($\alpha$) to evolve does not qualitatively change our results. (A) The color maps show the average fraction of parental lineages using each adaptation strategy (epigenetic switching, ES; bistable adaptation, BA; genetic adaptation, GA) of 3 evolutionary replicas under the corresponding mutation step-size ($M$) and environmental fluctuation frequency ($\nu$). Each simulation ran for 10000 generations with evolutionary parameters $N = 10000, s_t = 6, u = 0.03$ and $k = 80, n_H = 6, K_D = 45$ and $a = 0.25$ as the initial genotype ($\theta_1$). The corresponding (B) average bistable fraction ($\langle f_B \rangle_{\text{sim}}$) and (C) average $\alpha$ ($\langle \alpha \rangle_{\text{sim}}$). For some examples, the dynamics over time for the geometric mean fitness per cycle ($W_{\text{cycle}}$), the average basal activity ($\langle \alpha \rangle_{\text{cycle}}$), and the average bistable fraction ($\langle f_B \rangle_{\text{cycle}}$) per cycle, as well as the fraction of parental lineages using ES as the adaptation strategy per cycle, are shown: (D) $\nu = 0.1$ and $M = 5$; (E) $\nu = 0.02$ and $M = 2.6$; (F) $\nu = 0.002$ and $M = 5$; and (G) $\nu = 0.01$ and $M = 5$.

5.12 Population fitness depends on environmental fluctuation frequency and reflects a trade-off between adaptation time and phenotypic robustness. (A) Hill coefficient $\langle n_H \rangle_{\text{sim}}$ averaged over ten independent replica simulations for mutation step-size ($M$) and environmental fluctuation frequency ($\nu$). (B) Average geometric mean fitness per cycle $\langle W_{\text{cycle}} \rangle_{\text{sim}}$ for the same simulations. (C) Average population fitness at the second generation $\langle w \rangle_{2g}$ and (D) ninth generation $\langle w \rangle_{9g}$ after an environmental transition. Evolutionary strategies with a faster adaptation time had larger fitness after the transition (C) whereas those with robust phenotypes tended to have a larger fitness once they adapted to the new environment (D). All simulations ran for 10,000 generations with identical evolutionary parameters and initial genotypes as in Figure 5.6.

5.13 Population fitness trade-off is not observed in the absence of biochemical noise. Same simulation parameters as Figure 5.12 but with deterministic gene expression (i.e. CONTROL simulations). (A) Hill coefficient $\langle n_H \rangle_{\text{sim}}$ averaged over ten independent replica simulations for mutation step-size ($M$) and environmental fluctuation frequency ($\nu$). (B) Average geometric mean fitness per cycle $\langle W_{\text{cycle}} \rangle_{\text{sim}}$ for the same simulations. (C) Average population fitness at the second generation $\langle w \rangle_{2g}$ and (D) ninth generation $\langle w \rangle_{9g}$ after an environmental transition. (E-F) Two examples showing that the lower average $n_H$ values observed in the simulations with $M = 5$ arise from a lack of constrain over the actual nonlinearity value ($\nu = 0.1, \nu = 0.01$).
5.14 Increasing the mutation rate makes population fitness sensitive to mutation step-size ($M$) and affects qualitative trends in adaptation time and phenotypic robustness. (A,E) Average population Hill coefficient $\langle n_H \rangle_{\text{sim}}$ of 3 evolutionary replicas for mutation step-size ($M$) and environmental fluctuation frequency ($\nu$). (B,F) Average geometric mean fitness per cycle $\langle W_{\text{cycle}} \rangle_{\text{sim}}$ for the same simulations. (C,G) Average population fitness at the second generation $\langle w \rangle_{+2g}$ and (D,H) ninth generation $\langle w \rangle_{+9g}$ after an environmental transition. All simulations ran for 10,000 generations with identical evolutionary parameters and initial genotypes as in Figures 5.6 and 5.12, except for mutation rates ($u$) listed above each plot.

5.15 Increasing the selection pressure increases the population fitness without affecting the qualitative trends in adaptation time and phenotypic robustness. (A,E) Average population Hill coefficient $\langle n_H \rangle_{\text{sim}}$ of 3 evolutionary replicas for mutation step-size ($M$) and environmental fluctuation frequency ($\nu$). (B,F) Average geometric mean fitness per cycle $\langle W_{\text{cycle}} \rangle_{\text{sim}}$ for the same simulations. (C,G) Average population fitness at the second generation $\langle w \rangle_{+2g}$ and (D,H) ninth generation $\langle w \rangle_{+9g}$ after an environmental transition. All simulations ran for 10,000 generations with identical evolutionary parameters and initial genotypes as in Figures 5.6 and 5.12, except for selection pressure ($s_t$) listed above each plot.

5.16 Decreasing the population size increases the variation between independent replicas, without affecting the qualitative trends in fitness, adaptation time and phenotypic robustness. (A,E) Average population Hill coefficient $\langle n_H \rangle_{\text{sim}}$ of 3 evolutionary replicas for mutation step-size ($M$) and environmental fluctuation frequency ($\nu$). (B,F) Average geometric mean fitness per cycle $\langle W_{\text{cycle}} \rangle_{\text{sim}}$ for the same simulations. (C,G) Average population fitness at the second generation $\langle w \rangle_{+2g}$ and (D,H) ninth generation $\langle w \rangle_{+9g}$ after an environmental transition. All simulations ran for 10,000 generations with identical evolutionary parameters and initial genotypes as in Figures 5.6 and 5.12, except for population size ($N$) listed above each plot.
5.17 **Increasing $n_H$ increases the average population fitness in both environments.** (A) Contour plots as a function of biophysical parameters with fixed $k = 80$ where the steady state ($A^*$) for LOW are $A^{(L)} \pm 1\%$ (dark green) and for HIGH are $A^{(H)} \pm 1\%$ (light green). (B) I analytically calculated the average fitness of an infinite, clonal population with stationary protein distribution $\rho(A)$ given $n_H$, $K_D$ and fixed $k = 80$. The normalized distribution $\rho(A)$ was estimated numerically for each set of biophysical parameters (see Section 3.3.1). I calculated the expected fitness in each environment by integration, $E(w^{(E)}) = \sum_a w^{(E)}(a) \cdot \rho(a)$. 

5.18 **Increasing $n_H$ decreases the rate of epimutations.** (A) Effect of the Hill coefficient $n_H$ value over the stationary distribution of protein number ($\rho(A)$) with $k = 80$ and $K_D = 45$ fixed. The monostable cases are shown as dotted lines. (B) The stochastic switching rate from the LOW to HIGH phenotypic states ($1/\tau_{L\rightarrow H}$, light green) and from HIGH to LOW phenotypic states ($1/\tau_{H\rightarrow L}$, dark green) as the Hill coefficient $n_H$ increases. Only the bistable cases are shown (i.e. $n_H \geq 4$).
List of Abbreviations and Symbols

Symbols

\[ x \leftarrow y \quad \text{assign } y \text{ value to } x \text{ variable} \]
\[ \langle \bullet \rangle \quad \text{average} \]
\[ P(x|y) \quad \text{conditional probability of } x \text{ given } y \]
\[ \mathbb{E}(\bullet) \quad \text{expected value} \]
\[ e^x \quad \text{exponential of } x \]
\[ \exp(\bullet) \quad \text{exponential} \]
\[ x! \quad \text{factorial of } x \]
\[ \forall \quad \text{for all} \]
\[ \iff \quad \text{if and only if} \]
\[ \in \quad \text{in} \]
\[ \inf\{\bullet\} \quad \text{infimum} \]
\[ dt \quad \text{infinitesimal quantity of time } t \]
\[ \min(\bullet) \quad \text{minimum} \]
\[ P(\bullet) \quad \text{probability} \]
\[ \propto \quad \text{proportional to} \]
\[ \{\bullet\} \quad \text{set of elements} \]
\[ \mathbb{R} \quad \text{set of real numbers} \]
\[ \mathbb{R}_+ \quad \text{set of positive real numbers} \]
\[ \mathbb{R}_{\geq 0} \quad \text{set of non-negative real numbers} \]
\[ X^T \] transpose of a matrix \( X \)

\[ x \sim \mathcal{N}(\mu, \sigma^2) \] \( x \) is distributed normally with mean \( \mu \) and variance \( \sigma^2 \)

\[ x \sim \mathcal{U}(a, b) \] \( x \) is uniformly distributed between \( a \) and \( b \)

Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>AIC</td>
<td>Akaike Information Criterion</td>
</tr>
<tr>
<td>BA</td>
<td>Bistable Adaptation</td>
</tr>
<tr>
<td>BIC</td>
<td>Bayesian Information Criterion</td>
</tr>
<tr>
<td>CM</td>
<td>common condition medium</td>
</tr>
<tr>
<td>CME</td>
<td>Chemical Master Equation</td>
</tr>
<tr>
<td>CV</td>
<td>Coefficient of Variation</td>
</tr>
<tr>
<td>DCME</td>
<td>Delay Chemical Master Equation</td>
</tr>
<tr>
<td>DNA</td>
<td>Deoxyribonucleic Acid</td>
</tr>
<tr>
<td>DIC</td>
<td>Deviance Information Criterion</td>
</tr>
<tr>
<td>EE</td>
<td>Enhancer Element</td>
</tr>
<tr>
<td>eRNA</td>
<td>enhancer RNA</td>
</tr>
<tr>
<td>ES</td>
<td>Epigenetic Switching</td>
</tr>
<tr>
<td>( f_B )</td>
<td>fraction of bistable individuals in a population</td>
</tr>
<tr>
<td>fL</td>
<td>femtolitre</td>
</tr>
<tr>
<td>FSP</td>
<td>Finite State Projection algorithm</td>
</tr>
<tr>
<td>GA</td>
<td>Genetic Adaptation</td>
</tr>
<tr>
<td>IEG</td>
<td>Immediate Early Gene</td>
</tr>
<tr>
<td>iid</td>
<td>independent and identically distributed</td>
</tr>
<tr>
<td>KCL</td>
<td>Potassium Chloride</td>
</tr>
<tr>
<td>MCMC</td>
<td>Markov Chain Monte Carlo</td>
</tr>
<tr>
<td>min</td>
<td>minutes</td>
</tr>
<tr>
<td>Acronym</td>
<td>Full Form</td>
</tr>
<tr>
<td>----------</td>
<td>------------------------------------------</td>
</tr>
<tr>
<td>mRNA</td>
<td>messenger RNA</td>
</tr>
<tr>
<td>MRW</td>
<td>Metropolis Random Walk algorithm</td>
</tr>
<tr>
<td>nM</td>
<td>nanoMolar</td>
</tr>
<tr>
<td>ODE</td>
<td>Ordinary Differential Equation</td>
</tr>
<tr>
<td>PCR</td>
<td>Polymerase Chain Reaction</td>
</tr>
<tr>
<td>Pol II</td>
<td>RNA Polymerase II</td>
</tr>
<tr>
<td>RNA</td>
<td>Ribonucleic Acid</td>
</tr>
<tr>
<td>RNA Pol</td>
<td>RNA Polymerase</td>
</tr>
<tr>
<td>smFISH</td>
<td>single molecule Fluorescent In Situ Hybridization</td>
</tr>
<tr>
<td>TF</td>
<td>Transcription Factor</td>
</tr>
<tr>
<td>TS</td>
<td>Transcription Site</td>
</tr>
</tbody>
</table>
Acknowledgements

I want to thank my advisor and mentor, Professor Nicolas Buchler, for his constant support and guidance. During the time working in his laboratory, I have been able to freely explore and develop my scientific interests, always motivated by his great knowledge and constant inquiries. Particularly, I want to thank Prof. Buchler for his patience towards my obsessions with mathematical and evolutionary jargon, and his relentless dedication for helping me to improve my academic writing and presentation skills.

I would also like to thank all the other members of my thesis committee: Prof. Ryan Baugh, Prof. Katia Koelle, and Prof. Josh Socolar, for their insightful comments and encouragement on my academic projects. Similarly, I am thankful to our collaborators, Prof. Anne West and Liang-Fu Chen, for the opportunity to work with them, and for their patience and persistence to teach me and learn from me.

The community support has also been key for my success during these years, so I want to thank all the members of the Buchler lab for the interesting discussions, as well as my PhD program fellows for sharing this experience with me. In particular, I want to thank Liz Labriola for assisting me all this time and being there for me.

Finally, I want to thank all my friends and family, who have given me the strength to pursue my dreams. I thank my mom for always picking up the phone and listening to me, on the good days and the bad ones. I thank my dad for sharing my excitement for my research and for his unconditional support. Specially, I want to thank Rotem
Ben-Shachar, Brad Moore and Edgar Medina for making Grad School such a fun and wonderful experience. And more than anyone, I want to thank Sur Herrera for sharing the successes and the failures, for supporting me at every step, and for being my biggest motivation to improve myself every day.
1 Introduction

1.1 Overview

Organisms live in a highly dynamic environment, and they must be able to keep pace with frequent changes in order to survive. The environmental fluctuations vary widely in their temporal scale—from seconds to centuries—as well as in their form—from abrupt pulses to gradual changes—and their effect on the organisms—from beneficial to lethal. Consequently, evolution has given rise to a wide variety of mechanisms to adapt to these fluctuations. For instance, single cells have evolved to sense and biochemically respond to short-term changes in their surroundings, e.g. producing specific molecular machinery in the presence of certain nutrients (Ozbudak et al., 2004; Venturelli et al., 2015) or activating a stress response in extreme conditions (Dragosits et al., 2014). These responses are achieved through the precise regulation of the time and level of expression of the many genes that comprise an organism’s genotype—i.e. the genetic information encoded in the DNA sequence or genome of an organism. When the environmental changes persist for a long time, populations can adapt through adjustments that are inherited from generation to generation. For example, the genome is transmitted from parents to offsprings every generation,
and any changes in the genotype –DNA sequence mutation– can be inherited. This phenomenon is called genetic adaptation. For example, bacteria can mutate a gene which now confers resistance to an antibiotic and this mutation will be inherited by all its descendants (Palmer and Kishony, 2013). Interestingly, the phenotype –i.e. the gene expression state– can also be inherited through generations without requiring an underlying genetic mutation. This phenomenon is called epigenetic inheritance, and represents a form of cellular “memory” (Henikoff and Greally, 2016). Examples of epigenetic inheritance include diverse cell types in a human body, which have the same genotype, yet these cell types can display radically different phenotypes (e.g. gene expression patterns, morphology, and functions). Moreover, these epigenetic states are heritable because the division or reproduction of a committed cell type (e.g. muscle cell) results in the same type in descendant cells (Wang et al., 2009).

How organisms deal with the environmental fluctuations is also affected by the noise involved in the cellular response. All biochemical events are inevitably noisy because, among other reasons, many molecular species are present in low numbers and the reactions are random probabilistic events (i.e. stochastic) (Rao et al., 2002). How can organisms accurately sense the environment and induce the appropriate response in the presence of this ubiquitous noise? Interestingly, organisms have developed mechanisms not only to control but to exploit this biochemical noise (Raser and O’Shea, 2005).

The purpose of this thesis is to gain insights into how organisms can use epigenetic inheritance and the stochasticity of biochemical processes to deal with a fluctuating environment. To accomplish this, I have explored two cases at different temporal and structural scales using mathematical models and computational experiments. In the first part of my thesis, I analyze the information available in sparse single cell measurements to better characterize the stochasticity of gene expression regulation. In the second part of my thesis, I study the evolutionary advantage of using
epigenetics to exploit the natural noise in gene expression when facing a fluctuating environment. Before describing these projects in the subsequent chapters, I will first introduce the biological concepts related to gene regulation, biochemical noise, epigenetic inheritance and evolution.

1.2 Gene expression regulation

*Genes* represent the information unit of life, and are encoded in the DNA sequence or *genome* of each organism. Each gene is composed by a DNA sequence encoding a particular molecule—functional RNA or protein—and the *regulatory information* for its adequate expression. Briefly, in order to *express* a protein-coding gene, its *coding sequence* in the DNA must be transcribed into a messenger RNA (mRNA) and this mRNA translated into a protein. *Transcription* and *translation* are part of the *central dogma of molecular biology*, which was initially proposed by Francis Crick in 1956 (Figure 1.1A). Potentially, each of these steps is subject to regulation, however most research has focused on the regulation of transcription because this is often a limiting step for gene expression. Here, I will focus on the regulatory mechanisms of transcription.

The molecular mechanisms involved in transcriptional regulation vary significantly between bacterial and eukaryotic cells (Figure 1.1B-C). The transcriptional activity is often regulated by mediating the “accessibility” of the underlying DNA sequence to the core transcriptional machinery. The RNA polymerase—a protein complex that produces the transcript (mRNA)—first needs to bind and assembly at the gene promoter before starting the process of transcription. A gene is called *constitutive* if its expression is constant or non-actively regulated. A gene might be repressed by one or multiple *repressor* proteins bound around the promoter region, which block the access or progress of the RNA polymerase. A gene might be activated by the binding of *activator* proteins that increase the accessibility or affinity
Figure 1.1: **Central dogma of molecular biology & transcriptional regulation.** (A) The central dogma of molecular biology states that the genetic information is encoded in the DNA sequence, which can copy itself – *replication* – or be transferred to RNA – *transcription* –, and then transferred to protein – *translation*. (B) A diagram of transcription regulation in bacteria. A constitutive gene is expressed without active regulation; the gene can also be regulated by the binding and unbinding of transcription factors, which can work as repressors blocking the binding of the RNA polymerase (RNA Pol), or as activators facilitating the binding of the RNA Pol, which transcribes the gene into messenger RNA (mRNA). (C) Analogously, transcription in eukaryote cells is regulated by transcription factors, but more components are involved, including the holoenzyme complex (orange/brown), enhancer elements, and chromatin structure regulators (not shown). Additionally, the gene sequence might contain intron sequences, which are removed after transcription, and only the exon sequences are translated.

of the RNA polymerase for the promoter. These repressor and activator proteins are known as transcription factors (TFs), and the regulation of their own expression and localization in the cell allows for complex regulation (see Section 1.4). In addition to TFs, other regulatory elements have been described in eukaryotic cells, such as enhancer sequences and chromatin structure regulators (Figure 1.1). An enhancer sequence is a TF binding sequence distal to the promoter that helps recruit the RNA polymerase complex (Ong and Corces, 2011). Chromatin is a complex structure composed primarily of histones, which bind and wrap DNA, and package large genomes in the nucleus of the cell, as well as regulate the replication and transcription of DNA sequence (Schreiber and Bernstein, 2002).

Gene expression is often described by the average change per unit of time (i.e.
assuming an ordinary differential equation, ODE) of mRNA \((m)\) and protein \((x)\) concentration in the cell:

\[
\frac{d}{dt}m = g_+ - g_-
\]

\[
\frac{d}{dt}x = h_+ - h_-
\]

where \(g_+, g_-\) correspond to the mRNA synthesis and degradation rates, and \(h_+, h_-\) to the protein synthesis and degradation rates, respectively. By the law of mass-action (Turner et al., 2004), both degradation rates \(g_-, h_-\) are functions of the concentration of the molecule being degraded \((m, x\) respectively). Other regulatory proteins might take part on the degradation process, and these might also depend on time or an external stimulus. However, I will consider the simplest scenario, where the major decrease on these molecules occurs by their dilution as the cell grows and divides (Rosenfeld et al., 2002). The degradation rates are linear functions:

\[
g_-(m) = \gamma_m \cdot m
\]

\[
h_-(x) = \gamma_x \cdot x
\]

where \(\gamma_m, \gamma_x \in \mathbb{R}_+\) are the degradation constants.

The mRNA synthesis rate depends on the transcriptional regulation. If a gene is expressed constitutively, the \(g_+\) is a constant rate:

\[
g_+ = k_m
\]

with \(k_m \in \mathbb{R}_+\). If the gene expression is regulated by a transcription factor \(Y\), the mRNA synthesis rate is often described as a Hill function:

\[
g_+(Y) = k_m \cdot \frac{Y^n_H}{K_D^n_H + Y^n_H}
\]

\[5\]
Figure 1.2: Regulation of transcription rates. Effect of the cooperativity ($n_H$) over the transcription rate function of the concentration of (A) an activator or (B) a repressor of the gene. The constant $K_D$ is the concentration at which the transcription rate is half of the maximum rate $k$, and the Hill coefficient $n_H$ quantifies the cooperativity effect of the transcriptional factors (Equation 1.6).

where the Hill coefficient $n_H \in \mathbb{R}$ defines the sigmoidicity of the function, and $K_D \in \mathbb{R}_+$ is the TF concentration at which the synthesis rate is half of the maximum synthesis rate $k_m \in \mathbb{R}_+$. The Hill coefficient $n_H$ has been traditionally associated with the number of TF binding sites (Hill, 1910), although this relation is not strictly true. Even if a high $n_H$ implies a high level of cooperativity (i.e. an initial binding event affecting the rate of following bindings), other biochemical details of the DNA-TF interaction cannot be deduced from the Hill function parameter (Santillán, 2008). Many gene regulation systems often exhibit a sigmoidal relationship, making this phenomenological equation extremely useful to characterize and model gene regulation. The sign of the Hill coefficient determines the “direction” of the TF regulation: if $n_H$ is positive, the protein $Y$ acts as an activator (Figure 1.2A); if $n_H$ is negative, then $Y$ is a repressor (Figure 1.2B).

As mentioned previously, regulation often occurs at transcription. The protein synthesis rate due to translation of mRNA is a linear function of the mRNA concentration:

$$h_+(m) = k_x \cdot m$$

(1.7)
The timescale to reach steady state is determined by the underlying degradation rates, and mRNA is often degraded faster than protein. Thus, this two-level system can be further simplified by the *quasi-steady state assumption* because mRNA concentration often quickly reaches its steady state after any perturbation.

If \( g_-(m) = \gamma_m \cdot m \) and \( g_+ \) is independent of \( m \):

\[
\frac{d}{dt} m = g_+ - (\gamma_m \cdot m) = 0 \iff m = \frac{g_+}{\gamma_m}; \tag{1.8}
\]

and then:

\[
\frac{d}{dt} x = h_+(m) - h_-(x) \approx h_+ \left( \frac{g_+}{\gamma_m} \right) - h_-(x); \tag{1.9}
\]

This assumption is equivalent to ignoring the time delay between the mRNA concentration adjusting after a perturbation (e.g. a transcription factor concentration increases and stimulates the mRNA synthesis rate). Under this assumption, we condense Eqs. 1.1 and 1.2 into one equation:

\[
\frac{dX}{dt} = f_+ - f_-(X) \tag{1.10}
\]

where the synthesis \( f_+ \) can be a function of a transcription factor \( Y \), and \( f_-(X) \) is usually represented as the linear function \( \gamma \cdot X \), with \( \gamma \in \mathbb{R}_+ \).

1.2.1 *Deterministic dynamics of gene expression*

Gene expression dynamics are dictated by Eq. 1.10, which can be solved analytically—if possible— or numerically (e.g. Euler method). Gene expression is said to have reached a *steady state* \( (X^*) \) whenever the synthesis and degradation rate are equal, and there is no net change on the expression level (i.e. \( dX^*/dt = 0 \)). This steady state is called *stable* if it is an *attractor*, i.e. the system is attracted back to the same steady state after an arbitrary small perturbation \( \epsilon \in \mathbb{R}_+ \):

\[
\frac{d(X^* + \epsilon)}{dt} < 0 \tag{1.11}
\]
Figure 1.3: Steady state of gene expression and its stability. A steady state \( X^* \) occurs whenever the synthesis and degradation rate are equal, i.e. \( dX^*/dt = 0 \). (A) An example with a stable steady state: if a perturbation around the steady state decreasing the gene expression \( X^*-\epsilon \) results in \( d(X^*-\epsilon)/dt > 0 \), and a perturbation increasing the gene expression level \( X^*+\epsilon \) results in \( d(X^*+\epsilon)/dt < 0 \); i.e., the steady state works as an attractor. (B) An example showing both a stable and an unstable steady state. The unstable steady state works as a repeller, i.e. a perturbation around this steady state is not attracted back.

\[
\frac{d(X^*-\epsilon)}{dt} > 0
\]  

(Figure 1.3). Otherwise, the steady state is called unstable, and it is a repeller.

1.2.2 Stochastic dynamics of gene expression

Representing the gene expression dynamics as ODEs is valid description of the mean concentration in the limit of large numbers of molecules and large volumes. However, ODE systems assume both deterministic dynamics and continuous change over the gene expression \( X \). Nevertheless, all biochemical events are inherently stochastic because they are probabilistic events that depend on the collision and interaction of individual molecules (e.g. mRNA and proteins are discrete molecules) (Turner et al., 2004). The deterministic representation of mean concentrations (i.e. ODE system) is in general a good approximation if a large number of independent events occur (e.g. across a population of cells). In biology, this requirement is rarely satisfied. Often, the number of molecules involved in biochemical reactions are very low, e.g. genes are present in low copy numbers (e.g. 1-2) in most cell types. And given the discreteness
of the regulatory molecules, the effect of random fluctuation is relatively higher when molecules are at low numbers. The coefficient of variation (CV) is often used to quantify the relative noise in a system, and it is defined as the ratio of the standard deviation $\sigma_X$ over the mean $\mu_X$. Particularly, in a Poisson process with rate $\lambda$ the CV is equal to $1/\sqrt{\lambda}$. Then, in a simple example with constant synthesis ($f_+ = k$) and linear degradation ($f_-(X) = \gamma \cdot X$), the expected CV is $1/\sqrt{k + \gamma \cdot X}$, given that the next event is described as a Poisson process. Consequently, the relative noise in the system is expected to increase as the number of molecules involved decreases.

When the inherent stochasticity of biochemical reactions is considered, the gene expression is better modeled as individual probabilistic events that either increase –synthesis– or decrease –degradation– the number of proteins $X$ by 1:

$$X \xrightarrow{f_+} X + 1$$  \hspace{1cm} (1.13)

$$X \xrightarrow{f_-} X - 1$$  \hspace{1cm} (1.14)

where $f_+, f_-$ are known as the *propensities* of the reactions. These propensities are analogous to the rates described in the deterministic context (Eq. 1.10); but instead of a continuous “flux”, $f_+ \, dt$ and $f_- \, dt$ correspond to the probability that the synthesis and degradation reactions, respectively, will occur within an infinitesimal time interval $dt$, given the current number of proteins $X$ (Figure 1.4A). The gene expression dynamics are described by a random walk through the system possible states (i.e. the gene expression level or number of proteins), where the waiting time between reactions is assumed to decay exponentially, and the probability of “jumping” between states depends on the reaction propensities and the current state (Figure 1.4B). This process can be simulated using the Gillespie algorithm (Gillespie, 1977).

Importantly, the concept of deterministic steady state needs to be refined in this stochastic perspective because the stochastic biochemical reactions will continuously
perturb the system. Gene expression is better represented by the probability distribution of displaying a particular expression level \( X \) (i.e. a state) at a given time \( t \), \( P(X, t) \). The change over time of this probability distribution is described by the Chemical Master Equation (CME), which is the ensemble of ordinary differential equations \( \partial P(X, t)/\partial t \) for each possible state \( X \). In the simple gene expression model presented above (Eqs. 1.13-1.14), the CME is:

\[
\begin{align*}
\frac{\partial P(X - 1, t)}{\partial t} &= f_+(X - 2) \cdot P(X - 2, t) + f_-(X) \cdot P(X, t) \\
&\quad - (f_+(X - 1) + f_-(X - 1)) \cdot P(X - 1, t) \\
\frac{\partial P(X, t)}{\partial t} &= f_+(X - 1) \cdot P(X - 1, t) + f_-(X + 1) \cdot P(X + 1, t) \\
&\quad - (f_+(X) + f_-(X)) \cdot P(X, t) \\
\frac{\partial P(X + 1, t)}{\partial t} &= f_+(X) \cdot P(X, t) + f_-(X + 2) \cdot P(X + 2, t) \\
&\quad - (f_+(X + 1) + f_-(X + 1)) \cdot P(X + 1, t)
\end{align*}
\]

When the change on the probability distribution is zero for all possible states (i.e. \( \partial P^*(X, t)/\partial t = 0 \ \forall \ X \)), the system is in a stationary state (Figure 1.4C). The stationary distribution represents the behavior of single cell over a long period of time, or the distribution of states in a population of cells at any given time. The stationary distribution can be obtained analytically or numerically (e.g. using the Gillespie algorithm).

The reaction propensities only depend on the current state of the system \( X \), which in more complicated systems might include several molecules (e.g. mRNA and transcription factors), but not directly on time. This implies that the probability
of next reaction behaves as a Poisson process, i.e. independent events occurring randomly in time with a constant probability $\lambda dt$. In the case presented above, the probability of a reaction occurring is the sum of all reaction probabilities, $\lambda dt = (f_+(X) + f_- (X)) dt$; and the distribution of the waiting time $\tau$ between reactions is exponential, $F(\tau) = 1 - e^{\lambda \tau}$.

1.3 Biochemical noise

Some mechanisms can help organisms to reduce or control the biochemical noise. For instance, increasing the number of molecules of the species involved in a particular biochemical reaction can reduce the relative effect of individual stochastic events (e.g. in a system with 1000 molecules, the expected CV is $\sim 0.03$, while in a system with only 10 molecules the $CV \approx 0.3$). More elaborate mechanisms also exist; for example, negative feedback regulation, where a gene represses its own expression, has also been shown to reduce the biochemical noise in a system (Becskei and Serrano, 2000; Alon, 2007). The suppression of molecular fluctuations is strongly limited because usually a series of biochemical events are involved in any function, and the
minimum coefficient of variation decreases with the root of the product of the number of molecules for each species involved (Lestas et al., 2010). It may be very expensive for organisms to be accurate in their cellular responses and this cost might constrain the evolution of these regulatory mechanisms.

On the other hand, many organisms may take advantage of this intrinsic stochasticity of biochemical systems (Eldar and Elowitz, 2010). One functional advantage of biochemical noise is the probabilistic differentiation of a genetically homogeneous population, where the underlying gene regulatory system is triggered by noise and drives the system to distinct stable expression states or phenotypes. This has been observed both in multicellular organisms and microorganisms. In general, this strategy allows for the distribution of states or functions among the population in an efficient manner, that might be otherwise difficult or impossible by a deterministic program (Losick and Desplan, 2008).

Importantly, the level of biochemical noise can be tuned by natural selection. For example, the bias or ratio of probabilistic differentiation is far from random and significantly different among species (Losick and Desplan, 2008); and it has been observed that different environmental conditions can shape this ratio (Wang et al., 2015). The specific proportion of the population in each state will depend on the underlying genetic system and its biophysical parameters. To be able to characterize the role and control of noise, as well as its evolutionary constraints, we need to understand how the gene circuit dynamics work and their emergent properties.

1.4 Gene circuits and emergent properties

Most cellular functions require the coordinate action of multiple genes. As mentioned above, the appropriate regulation of the level and timing of the gene expression is necessary for the proper function of any organism. A gene circuit is the set of transcription factors interactions that regulates a specific cellular function, such as
processing information from the environment or choosing a particular cell fate (Kim et al., 2009). These gene networks tend to be very complex, involving many genes and many interactions (Tyson and Novák, 2010). Interestingly, certain recurrent patterns or motifs have been detected in these complex networks, and these motifs have been associated with specific functions in the regulatory programs (Milo et al., 2002; Alon, 2006). Some well characterized examples include positive and negative feedback loops, and coherent and incoherent feedforward loops (Figure 1.5).

The recurrence of these regulatory units or motifs is not a coincidence. Due to physical and mathematical constraints, certain regulatory circuits give rise to specific dynamic properties. These dynamic properties are called emergent properties of the gene regulatory circuits. For example, the negative feedback loop, where a gene represses its own synthesis (Figure 1.5), can be used to accelerate the response time of a gene (Alon, 2007). In the presence of long time delays the negative feedback loop can become unstable and exhibit oscillations (Ferrell et al., 2011). Analogously, some specific feedforward loops have been associated to particular cellular functions.
The *coherent feedforward loop* type 1 can delay the activation (deactivation) after a stimulus without affecting the deactivation (activation) time when using an AND (OR) logic in the co-regulation of the $Z$ gene (Figure 1.5). This type of motif can help filter transient pulses of stimulus (loss of stimulus). The *incoherent feedforward loop* type 1 can also allow for fast responses in a similar way to the negative feedback, or reach biochemical adaptation –i.e. a transient response in the presence of a constant stimulus– if the repression of $Z$ by the TF $Y$ is strong enough (Ma et al., 2009).

**1.4.1 Bistability**

*Positive feedback loops* in gene circuits give rise to *bistability*. A system is bistable (or multistable) if it has two (or more) distinct stable states. The simplest gene circuit capable of displaying bistability is a self-activated gene (Figure 1.6A):

\[
\frac{dX}{dt} = f_+(X) - f_-(X)
\] (1.16)

\[
f(X) = k \cdot (\alpha + (1 - \alpha) \frac{X^n_H}{X^n_H + K^n_D})
\] (1.17)

where the parameters are mostly the same as described in Eq. 1.6. The basal activity term $\alpha$ is added to allow for the first steady state to be distinct of zero (Figure 1.6B). The positive feedback is necessary for bistability to arise, but not enough (Thomas and Kaufman, 2001); only some regulatory parameters will generate bistability. For instance, a high nonlinearity is required for two or more stable steady states to coexist. This nonlinearity can arise from multiple TF molecules binding to the promoter (i.e. cooperativity) (Santillán, 2008).

**1.4.2 Bistable switches**

As explained in Section 1.2.2, when the effect of biochemical noise is considered, the biochemical reactions are probabilistic events over time and the CME is a better
representation of the gene expression dynamics. In bistable systems, the *stationary distribution* of the corresponding CME (i.e. $\partial P^*(X,t)/\partial t = 0, \forall X$) is a bimodal *stationary distribution*, where each mode corresponds to a steady state, and the valley between those is related to the probability of transition between states (i.e. the probability of moving from one mode to the other, Figure 1.7).

A useful analogy to understand the dynamics of a bistable system in the presence of biochemical noise is the *potential energy landscape* $V(X)$. Traditionally, this has been used to describe the probability or frequency of crossing an energy barrier in a chemical reaction by thermal noise or Brownian motion. This movement is described by the Langevin equation:

$$\lambda_B \frac{dX}{dt} = -\frac{\partial V}{\partial X} + \eta(t) \quad (1.18)$$

where $\lambda_B$ is the drag coefficient of the particle, and $\eta(t)$ is the noise term determined by the temperature, the Boltzmann’s constant, and $\lambda_B$. The potential energy is defined by $\partial V/\partial X = -\partial X/\partial t$. Then, the system is attracted to the local minima in
The potential energy landscape $V(X)$, and having a transition between attractors is equivalent to thermal activation over the energy barrier, i.e. the hill of the potential energy landscape. The probability of transitioning between stable states, i.e. moving over the activation energy barrier, depends on the slope of the potential energy $V(X)$ and the thermal noise, which in the case of Brownian motion is determined by the temperature and the drag coefficient of the particle (Bialek, 2000). Natural selection can tune these regulatory parameters to exhibit a specific rate stochastic transitions and population distribution.
1.5 Genetic vs epigenetic inheritance

*Genetic inheritance* is the transmission of the parental DNA sequence to their offspring. Classic evolutionary biology is founded on this Mendelian model of inheritance. However, it is known that evolution occurs at multiple levels, including non-genetic transmission, such as epigenetic regulation, and behavioral learning (Jablonka and Lamb, 2005). It is clear that mother cells transmit much more than their DNA sequence to the daughter cells, and multiple examples have shown that non-genetic information can define the daughter cell state (Veening et al., 2008) and directly affect the speed and direction of evolution (Day and Bonduriansky, 2011). The interplay of the genetic and non-genetic inheritance mechanisms and their general relevance on the evolutionary process are still being explored.

The term *epigenetics* was first defined by Waddington (1942) as the study of the causal mechanisms that give rise to an observed cellular state given the underlying genetic information. Waddington focused on the developmental process of multicellular organisms, where embryonic cells with exactly the same DNA sequence differentiate into multiple functional states to build a mature organism. Later, this definition was used to describe the molecular systems regulating the expression of “genetically determined potentialities” (Nanney, 1958). Eventually, the concept of *epigenetic inheritance* arose to explain the existence of “induced heritable changes” that perpetuate a particular pattern of gene expression even after the removal of the stimulus, e.g. lactose regulatory system in *Escherichia coli* (Novick and Weiner, 1957).

In general terms, *epigenetic switches* are systems capable of sustaining two heritable gene expression states without underlying DNA mutation (Iliopoulos et al., 2009). As mentioned above, the coexistence of these steady states requires bistability in the regulatory system (see Section 1.4.2). For example, an epigenetic switch can
emerge through a simple positive feedback loop in gene networks. As mentioned in section 1.4.2, the transition between alternative states can occur spontaneously due to stochastic fluctuations in gene expression (Figure 1.8). The rate of transitions will depend on specific regulatory parameters, and a particular phenotypic state could persist through multiple generations.

External stimulus can also regulate these epigenetic switches, controlling the state displayed by the cell, as well as the probability of transition to the alternative state. In absence of biochemical noise, the system will display hysteresis, i.e. the level of stimulus required to induce the system is different from the threshold to uninduce the system (Figure 1.9), and consequently the current state of the cell depends completely on its history. This phenomenon is known as cellular memory. When biochemical noise is considered, the external stimulus can affect the stochastic transition rate between states by changing the potential energy landscape (i.e. the energy barrier between states), and eventually create or destroy attractors in the system (i.e. from monostable to bistable and vice versa).

Several different epigenetic mechanisms are known at the molecular level. Dodd
et al. (2007) defines two major classes: the “cytoplasmic” class modulated by diffusible molecules, for example transcriptional factors, microRNAs, and prions; and the “chromosomal” class that involves DNA structure-associated molecules, including DNA methylation, and histone modifications. All of these mechanisms encompass a self-reinforcing feedback loop capable of switching between two or more stable phenotypes. The memory and regulation of each case can vary widely, from permanent transitions as in neuron differentiation in higher eukaryotes (Vaccarino et al., 2001), to labile transitions as in the case of bacterial persistence (Balaban et al., 2004).

Figure 1.9: **Induced transitions in epigenetic switches.** (A) Diagram of a bistable (epigenetic) switch dependent on a stimulus. One or two stable states exist in the system as a function of the stimulus level. In the bistable region two stable (thick line) and one unstable (dashed line) coexist. Overlooking the biochemical noise, the system is induced (green dotted line) only if the stimulus level is higher than $s_I$; but if the stimulus level is decreasing, the system “remembers” the previous state and only transitions to the uninduced state (red dotted line) if the stimulus is less than $s_U < s_I$. This shift between the induction and uninduction threshold ($s_I \neq s_U$) is known as **hysteresis**. (B) Diagram of the expression of the transcriptional factor of the epigenetic switch as the stimulus level changes over time, overlooking the biochemical noise. The potential energy function corresponding to different levels of stimulus, as well as the state of the system in these (blue dot), are shown as a reference. Notice that the system is bistable always that the stimulus $s$ is $s_U \leq s \leq s_I$, but the displayed state depends on the cellular history.
1.6 Fitness and adaptation in a static environment

The basic principle of evolution, as presented by Darwin and still accepted, is the “survival of the fittest”. Roughly, fitness is a measure of the probability of surviving –viability– and reproduction –fecundity– of an organism in a particular environment (Orr, 2009). Fitness is determined by the phenotype of the individual (i.e. set of observable characteristics or traits) that results from the interplay between the genotype (i.e. DNA sequence or genetic constitution of an individual) and the environment. Assigning absolute fitness is not an easy task because the components of fitness vary widely among taxa and conditions (e.g. division rate versus mating success). Nevertheless, the relative fitness –which is in general easier to determine– is often sufficient to analyze the evolutionary dynamics, given that selection is a differential process (winners vs. losers) (Orr, 2009).

Natural selection acts upon the differences of fitness among individuals in a population. Consequently, variation must exist in the population in order for the selection to act. Variation is introduced in the population by multiple mechanisms affecting either the genotype, including mutation, recombination, and horizontal gene transfer (Rosenberg, 2001), or just the phenotype (i.e. an epigenetic transition, or epimutation) (Rando and Verstrepen, 2007). Then, the population structure or distribution evolves by the appearance of variation and its selection, as well as genetic drift –the effect of stochasticity on the selection process– and migration. Adaptation corresponds to the “movement” or shift of the population phenotype distribution towards a higher fitness given the environment (Orr, 2005). The speed and path of adaptation depends directly on, in addition of the forces just described, the shape of the fitness landscape –i.e. the map between genotypes and their corresponding fitness. Even though much is known about selection and mutation, very little is actually understood about the adaptation processes because the fitness landscape is often
unknown. Furthermore, the fitness landscape is expected to vary with the environment—abiotic elements—as well as with the population structure—biotic elements—(Meyers and Bull, 2002). Particularly, there are two important properties associated with the fitness landscape: robustness and evolvability. A population display robustness if low phenotypic variation is observed after a genetic or environmental perturbation (Félix and Barkoulas, 2015). On the other hand, a population has high evolvability if heritable phenotypic variants can be generated, either by mutations or epimutations (Masel and Trotter, 2010).

1.7 Adaptation in fluctuating environments

The environment has many dimensions—e.g. temperature, chemicals, nutrients, predators, and competition—, each potentially varying in different forms and temporal scales. Adaptation process in fluctuating environments is an old question in evolutionary biology (Kimura, 1954; Cohen, 1966; Gillespie, 1972). The effect of environmental fluctuations on the adaptation process is easily understood when these two processes—fluctuations and adaptation—occur on different temporal scales. For instance, if the environment fluctuates infrequently, the fitness landscape is effectively static and the adaptation process becomes a problem of hill climbing optimization of the phenotype. On the other hand, if the environment fluctuates very quickly, the time average fitness effect might be a good representative of selective pressure on a particular phenotypic variant. Nevertheless, Cvijović et al. (2015) have shown that fluctuating environments can dramatically change the outcome of the adaptation process, sometimes with very unintuitive results, when these have similar temporal scales. For example, variants show differential fitness values between environments but are strongly deleterious on average can have a higher probability of being selected in the population than variants that are neutral or beneficial in both environments.

Consequently, exploring the relation between environmental fluctuation frequency
and the adaptation strategy employed by the organism has become a very active area of research (e.g. Venturelli et al. (2015); Mayer et al. (2016); Cerulus et al. (2016)). To date, the proposed strategies employed by organisms to deal with fluctuating environments can be roughly classified in three categories: sensing, mutating, or anticipating. These strategies are not necessarily mutually exclusive.

1.7.1 Sensing and responding to a stimulus

Many organisms have evolved biochemical networks to directly sense environmental cues and induce the appropriate phenotype for the given environment. This phenomenon is called adaptive phenotypic plasticity. This strategy requires the existence of reliable cues that are correlated with the changing environment, as well as the “machinery” encoded in the genotype to be able to sense these cues and efficiently induce the correct phenotype. Phenotypic plasticity enables an immediate adaptation to a specific environmental change. However, “ideal” plasticity is rarely observed in nature, which suggests that its evolution might be constrained by the costs and limitations of implementing phenotypic plasticity (Murren et al., 2015). The costs of plasticity include the maintenance –e.g. sensing machinery–, production –e.g. response machinery–, and regulation –e.g. accuracy and noise effects. Other limitations include extracting meaningful and trustful information from the environment and having a short response delay or lag (DeWitt et al., 1998). These properties, the specific environmental dynamics and the penalty of being “maladapted” determine the advantage of the adaptive phenotypic plasticity (Jablonka et al., 1995).

1.7.2 Adapting through genetic mutations

Genetic adaptation has for long been identified as the standard mechanism of evolutionary adaptation. Genetic adaptation takes place through the natural selection of genetic variants with different phenotypes and fitness. The genetic variants or
mutants might emerge randomly every generation –*de novo* mutations–, or being already present in the population –standing variation–, either because these mutants were neutral, or have not been purified by selection yet (Peter et al., 2012).

In contrast with phenotypic plasticity strategies, genetic adaptation has the advantage of generating potentially many different phenotypes, without requiring any previous knowledge of the environment. Nevertheless, if environmental fluctuations occur too fast or if the populations are too small, the spontaneous appearance of a beneficial mutation might be too infrequent. Some organisms may induce genetic variation to accelerate adaptation to a new environment (Sniegowski and Lenski, 1995; Perfeito et al., 2007; Rohner et al., 2013). Importantly, more deleterious than advantage mutations are expected, imposing a cost or *mutation load* in the population as the mutation rate increases (Levins, 1967).

1.7.3 *Anticipating the changes: Bet-hedging*

*Bet-hedging* is the phenomenon where the geometric mean of fitness of a genotype is maximized at the expense of the average fitness generally by the spontaneous generation of phenotypic variation in the population previous to the environmental change (Viney and Reece, 2013). The idea comes from the fact that in a fluctuating environment the geometric mean (i.e. the $n$th root of the product of the fitness during $n$ consecutive generations) becomes a better measurement of fitness than the arithmetic mean (i.e. the sum of the fitness during $n$ consecutive generations divided by $n$) because the former captures the multiplicative nature of fitness across generations and penalizes fitness variance in a population. For example, a population might face extinction in a short but extreme environment, resulting a zero fitness geometric mean, but could still display a high fitness arithmetic mean. Many examples of bet-hedging strategies exist along the three of life, including seeding times in plants (Childs et al., 2010) and bacterial persistence (Balaban et al., 2004).
Importantly, bet-hedging cases usually work by an underlying epigenetic switch (see Sections 1.4.2 and 1.5). This has two implications: (1) the genotype capable of displaying the multiple phenotypes via epigenetics must be previously selected, and (2) the level and type of variation generated in the population through epimutations can be tuned by natural selection. Bet-hedging via epigenetics has the cost associated with the random (non-induced) epimutations, but no sensing and responding machinery is required. If the phenotypes arising from epimutations are restricted to some specific states, bet-hedging can allow faster adaptation to the new environment compared to genetic adaptation.

1.8 Simulating evolution

It has long been recognized that given the inherent complexity and the long time scales of evolutionary processes, mathematical models are essential tools to study evolution (Servedio et al., 2014). By modelling the adaptation process, we can identify the diverse factors driving and constraining the evolutionary process, with the potential to systematically perturb or manipulate the conditions. Many of these scenarios might be difficult or even impossible to implement by experimentally (Murren et al., 2015). Additionally, as the complexity of the analyzed adaptation process increases, the complexity of the required mathematical model will increase and become analytically intractable. Then, performing computational simulations becomes a key instrument to investigate the adaptation process. By using a computer simulation to evolve organisms under a variety of conditions, we are able to study principles and laws of evolution at scales impossible by traditional experimental approaches.

The idea of simulating evolution was initially explored in 1950-60s by computer scientists with the aim of developing better optimization algorithms for engineering problems (Mitchell, 1998). Holland (1975) was the first to apply these algorithms in the theoretical framework of biological evolution as an abstraction of the adaptation
phenomena occurring in nature. Currently, many examples exist of computer simulations used to explore the adaptation process at the level of DNA sequence (Dutheil et al., 2009), gene circuits (Sorrells and Johnson, 2015), and populations (Lin et al., 2015).

Two fundamental models of evolution are generally applied in these simulations: *Wright-Fisher* and *Moran* models. The first one was originally developed to study the effect of a small finite population and random fluctuations on the adaptation process (i.e. *genetic drift*). The most important assumptions behind this model are that the generations do not overlap (i.e. the whole population reproduces and dies at exactly the same time) and a fixed population size. This model can be modified to include mutation and selection, among other relevant processes. The Moran model also assumes a fixed population size, but the birth-and-death events are simulated as stochastic events (Moran, 1958). The Moran model of evolution tends to be significantly more expensive computationally because the whole population needs to be updated and sorted at each life events, instead of just once at the end of the generation (i.e. \(N:1\) more times than the Wright-Fisher model with a population size \(N\)).

1.9 Organization of Dissertation

To gain insights into how organisms can use epigenetics and the intrinsic stochasticity of gene regulation to deal with a fluctuating environment, I have explored two distinct questions in this thesis: (1) how can the intrinsic stochasticity of gene expression be characterized from sparse single cell measurements; and (2) under which evolutionary conditions are stochastic epigenetic switches selected as the population strategy to deal with a fluctuating environment. Both of these questions required the use and development of stochastic models of gene expression, population dynamics, and evolution. Distinct approaches were used to address each of these questions,
one at the single cell, single stimulus scale, and another studying a population under a recurring environmental change. In Chapter 2, I develop a mathematical and statistical model to characterize the kinetics of a single cell, single gene behavior in response to a single environmental stimulus. I use a Bayesian inference approach to deduce the contribution of multiple chromatin states –epigenetics– of the gene promoter on the experimentally measured stochastic. In the following chapters, I study the evolutionary adaptation dynamics of epigenetic switches in a recurrent fluctuating environment, observing the evolution of gene regulatory circuit in a population under multiple environmental cycles. In Chapter 3, I present a mechanistic model of a self-regulatory gene where genetic adaptation as well as epigenetic switching can be implemented by the population, and how I analyze the evolutionary dynamics through computational experiments under a wide variety of conditions and model assumptions. In Chapter 4, I analyze the population dynamics during the adaptation process and the effect of the phenotypic landscape, taking advantage of the great detail obtained from these simulations –in contrast to most experimental approaches. And finally, in Chapter 5, I use the cells’ lineage information –i.e. their evolutionary history– to characterize the adaptation strategies selected in the population and their fitness advantage. A general conclusion and proposals for future directions are presented in Chapter 6.
Stochastic dynamics of early transcriptional response: Activity-dependent immediate-early gene expression in single neurons

Epigenetic mechanisms play an essential role in the development of both microbes and multicellular organisms (Dodd et al., 2007), by allowing cells to switch to an alternative phenotypic state without requiring any genetic mutations. This phenotypic state can be reversible or permanent depending on the regulatory parameters underlying the specific epigenetic switch, and the transitions between phenotypes can be spontaneous or triggered by an environmental cue (Iliopoulos et al., 2009). The intrinsic stochasticity of the system is fundamental for an epigenetic switch because it regulates the probability of transition and determines the stability of each phenotypic state. Thus understanding the underlying source of stochasticity is necessary for fully characterizing the behaviour of epigenetic switches.

Since the mid 1990s, important advances have been made in experimental techniques, allowing single molecule measurements within single cells under a wide variety of conditions. For example, the single-molecule fluorescence in situ hybridization
smFISH technique detects single mRNA molecules and active promoters by targeting a mRNA sequence of interest with fluorescent probes. Free mRNA molecules are individual bright points, while nascent mRNAs appear as bigger/brighter points because these accumulate around individual active promoters (Femino, 1998; Mueller et al., 2013). Important limitations exist for applying smFISH to many cell types, such as neurons. Studying gene expression in neurons requires primary cultures; and, given that mature neurons do not undergo cell division, the number of cells available for experiments is limited and expensive (Gordon et al., 2013). Additionally, if stochasticity is relevant in the cellular phenomena of interest, extracting information from these small samples ($n \approx 100$ cells) can be challenging. For these reasons, mathematical and statistical models are needed to determine the kinetic and stochastic properties of cellular systems, whether the organisms take advantage of the biochemical noise—e.g. bet-hedging strategies (Garcia-Bernardo and Dunlop, 2015; Cerulus et al., 2016)—or need to minimize noise effects—e.g. cell cycle commitment (Verdugo et al., 2013; Cardelli et al., 2016)—.

In this context, a particularly interesting case is the transcription of activity-dependent immediate-early genes (IEGs). The IEGs represent the first response to a wide variety of cellular stimuli, later coordinating a complex program of stimulus-specific gene expression. In neurons, the IEGs modulate the formation, maturation, and plasticity of neuronal synapses coupling extracellular stimuli with the corresponding intracellular adaptation (Leslie and Nedivi, 2011; Lyons and West, 2011). In the laboratory of Dr. Anne West (Neurobiology, Duke University; personal communication), smFISH measurements of IEGs mRNA expression in neurons have shown a significant cell-to-cell variation after induction of neural activity (e.g. exposing cells to potassium chloride, KCl). However, their measurements are restricted to small sample sizes, which has limited the characterization of the kinetics and intrinsic stochasticity of the activity-dependent immediate-early transcriptional response.
My goal is to develop a mathematical and statistical model that allows us to characterize the kinetics and intrinsic stochasticity of the activity-dependent immediate-early transcriptional response using small samples of single cell mRNA expression snapshots. This model will be trained using the IEGs smFISH measurements from the laboratory of Dr. Anne West.

In this chapter, I present a framework to study the stochastic dynamics of early transcriptional response (Section 2.1). I will briefly describe the role and relevance of activity-dependent IEGs in neurons. I will also review the known molecular mechanisms related to the IEG transcriptional regulation, using the Fos gene as an archetype. IEGs have attracted significant attention in the last thirty years not only for their activity in a broad range of cell types, but also the particular chromatin structure in the IEG promoter that allows for the immediate transcriptional response after stimulus. I will then explain how the single-molecule gene expression data was obtained, and I will highlight the challenges imposed by the small cell samples from primary neurons (Section 2.2). Afterwards, I will explain the model and algorithm technical details (Section 2.3). In order to include the intrinsic stochasticity of the molecular process –i.e. local biochemical events regulating the promoter activity–, and explore its contribution to cell-to-cell variability, I implemented a model of the transcriptional process, assuming each promoter allele is regulated independently and can be in one of two states –active or inactive–. To characterize the expected cell-to-cell variability given the model, the probability distribution of the gene expression is calculated using stochastic processes tools. Additionally, to evaluate the early transcriptional response, where the system has not reached a stationary state, the dynamics of the probability distribution after induction (i.e. the stimulus) are simulated using the Finite State Projection algorithm (FSP). Finally, I implemented a Bayesian Inference approach to obtain an estimate of those kinetic parameter val-
ues that best describe the experimental data. I also evaluated the confidence on these values estimates. This Bayesian approach allowed us to overcome the difficulty of the small and restricted experimental samples.

Using this algorithm, I evaluated the experimental data from West laboratory under alternative model assumptions corresponding to which molecular process were regulated by the induction stimulus. I show that the model is capable of recapitulating the system’s behavior only when the promoter activation rate is regulated by the stimulus (Section 2.4). Finally, I explored the effect of the sample size and the time points measured on the identifiability of the kinetic parameters using synthetic data. I show that the sample size needs to increase by one order of magnitude to have a significant improvement on parameter identifiability. The selection of the time points to make a measurement can have a considerable effect, without increasing the total number of cells being sampled (Section 2.5). The mathematical and statistical model developed here has allowed us to gain insights into the dynamical properties, as well as the role of stochasticity on this cellular process. Moreover, the framework is general enough to be applied to other dynamical systems with high stochasticity and limited sample sizes (Section 2.7).

2.1 Framework of stochastic dynamics of early transcriptional response

In our model, the gene promoter can display distinct discrete states, each characterized by a different synthesis rate. Here, I will focus on a 2-states model (Figure 2.1), where the promoter can be in (1) an inactive state ($\rho_{OFF}$) with a small –leaky– transcription level (synthesis rate $\mu_0$), and (2) an active state ($\rho_{ON}$) with a high transcription level (synthesis rate $\mu \gg \mu_0$). The transitions between promoter states are assumed to occur stochastically and independently in each promoter, with a promoter activation ($\rho_{OFF} \rightarrow \rho_{ON}$) rate $k_{ON}$ and a promoter deactivation ($\rho_{ON} \rightarrow \rho_{OFF}$) rate
Figure 2.1: Diagram of 2-states model of mRNA expression. Each cell has two copies of the gene, and each gene promoter (circles) can be either in active ($\rho_{ON}$) or inactive ($\rho_{OFF}$) state; an active promoter turns inactive with rate $k_{OFF}$, and an inactive promoter turns active with rate $k_{ON}$. Each promoter copy synthesizes mRNA molecules ($m$) with rate $\mu$ or $\mu_0$ if it is active or inactive, respectively; and the mRNA molecules are degraded with rate $\delta$. The stimulus ($s$) can potentially affect each of these molecular processes.

$k_{OFF}$. The number of copies of the promoter is assumed constant, e.g. two copies in a diploid cell ($[\rho_{ON}] + [\rho_{OFF}] = 2$, where $[*]$ represents the number of molecules per cell), and the total mRNA synthesis rate is $\mu[\rho_{ON}] + \mu_0[\rho_{OFF}]$ for any single cell. Additionally, the mRNA decays with a constant rate $\delta$.

This 2-states model can give rise to an interesting dynamic behavior known as transcriptional bursting, which has been proposed as a general property of gene expression. Transcriptional bursts are episodes of transcriptional activity followed by long periods of inactivity. This phenomenon has been observed in many organisms, including bacteria (Golding et al., 2005), yeast (Cai et al., 2008; Zenklusen et al., 2008), Drosophila (Fukaya et al., 2016) and mammals (Bahar Halpern et al., 2015). Additionally, significant theoretical work has been done to characterize the burst dynamics and its implications in cell-to-cell variability (i.e. gene expression noise) (Sanchez and Golding, 2013). In the simple model described here, the burst duration, i.e. the expected lifetime of the promoter active state, is regulated by the deactivation
**Figure 2.2**: Effect of parameters on gene expression and bursts dynamics in single cell. Diagram exemplifying the aspects of the dynamical system affected by each individual parameter in the model: $k_{ON}$ determines the frequency of burst, i.e. the expected time the promoter stays in the inactive state (e.g. blue line); $k_{OFF}$ determines the length of the bursts, i.e. the expected time the promoter stays in the active state (e.g. violet line); $\mu$ determines the magnitude of the burst, i.e. how much mRNA is accumulated once a promoter is in the active state (orange arrow); $\mu_0$ determines a basal mRNA signal (brown bar); and $\delta$ defines how the burst decays over time (gray arrow). The parameter values used are: $k_{ON} = 0.01 \text{ min}^{-1}, k_{OFF} = 0.02 \text{ min}^{-1}, \mu_0 = 0.1 \text{ mRNA/min}, \mu = 5 \text{ mRNA/min},$ and $\delta = 0.0462 \text{ min}^{-1}$. The graph shows the simulated dynamics of one cell, where $\rho_{1,2}$ are the individual promoter copies and the color reflects their state (green if active, $\rho_{ON}$; gray if inactive, $\rho_{OFF}$), and mRNA is the number of $m$ molecules over time. The average mRNA number in the particular simulation is also shown as a dashed yellow line.

rate ($k_{OFF}$), while the burst frequency, i.e. how often these burst events occur, depends on the activation rate ($k_{ON}$). The burst amplitude, i.e. the amount of mRNA produced in a burst event, depends on both the burst duration and the synthesis rate ($\mu$; Figure 2.2).

In the context of our model, a stimulus can increase expression levels by either increasing the synthesis rate ($\mu$), and/or the expected time the promoters are in active state (by either increasing $k_{ON}$ or decreasing $k_{OFF}$ values; Figures 2.1 and 2.2). Each of these mechanisms will display a distinct dynamic behavior (Figure 2.3). For instance, it has been shown that even if the mean expression level stays constant, differences in the burst dynamics can impact significantly the type of variation or noise in gene expression. For example, by slowing down the promoter transitions
Figure 2.3: Increasing gene expression average level. Examples of how the system dynamics are affected as individual parameters change. Same graphs and notation as Figure 2.2; the same parameters were used except for: (A) higher $k_{ON} \rightarrow 0.04 \text{ min}^{-1}$, (B) lower $k_{OFF} \rightarrow 0.005 \text{ min}^{-1}$, (C) higher $\mu \rightarrow 10 \text{ mRNA/min}$, and (D) higher $\mu_0 \rightarrow 1 \text{ mRNA/min}$. In all cases, the average mRNA levels increased with respect to Figure 2.2, but very different dynamics are observed.

(smaller $k_{ON}$, $k_{OFF}$ values) the mRNA distribution passes from a unimodal distribution with relatively low variation to a clear bimodal distribution, resulting in a large variation in gene expression (Munsky et al., 2012). Extracting the kinetic parameters $\{k_{ON}, k_{OFF}\}$ is essential for characterizing the dynamics and cell-to-cell variability associated with active-dependent gene expression. Even if the individual cell dynamics are not directly measured, the population distribution at any given time shows signatures of the underlying dynamic process (Figure 2.4). Then, temporal snap-
Figure 2.4: Population distributions display signatures of the gene expression dynamics. (A) A temporal snapshot (red line) of the population. (B) The stationary probability distribution of the population shown in (A). The corresponding biophysical parameters are shown in the bottom, and the expected average mRNA molecules number in the population (\(\langle \text{mRNA} \rangle \)) is 75. (C) Examples of how the stationary probability distributions change as individual parameters vary: \(k_{\text{ON}} \rightarrow 0.04 \text{ min}^{-1}, k_{\text{OFF}} \rightarrow 0.005 \text{ min}^{-1}, \mu \rightarrow 10 \text{ mRNA/min}, \) and \(\mu_0 \rightarrow 2.5 \text{ mRNA/min}.\) Notice that the expected average mRNA molecules number in the population is practically identical in all these cases (\(\langle \text{mRNA} \rangle \approx 150\)), while the populations show qualitatively different behavior (see Figure 2.2).

shots can be used to characterize the kinetic parameters associated to a particular gene expression model.

2.2 Activity-dependent, immediate-early transcriptional response

2.2.1 Transcriptional mechanisms of activity-dependent synapse development and plasticity

The brain is capable of converting sensory information –learning– into changes in neuronal function –response– in a highly plastic, adaptable manner. This process is
in part accomplished through complex activity-regulated transcriptional programs, where immediate-early gene transcription factors represent the first response, coupling the extracellular stimuli with intracellular adaptations (Lyons and West, 2011). A striking property of the IEGs, and the reason of their name, is the very fast initiation of their transcription after stimulus, which can be detected within one minute (Greenberg et al., 1986). This rapid response is possible given that the IEGs promoter is in a primed state before neural activity, with the RNA polymerase II (Pol II) complex and other transcriptional activators bound to the promoter (West and Greenberg, 2011) (see Section 2.2.2).

The functional response of the IEGs transcriptional activation depends on the cell type and the specific stimulus, which recruits distinct transcription factor complexes that target a particular set of downstream genes (Lyons and West, 2011). The associated complex patterns have raised the question of how the stimulus and IEGs achieve the required specificity in the cellular response; and considerable effort has been directed to identify the molecular mechanisms responsible. However, the extent to which stochasticity is associated with the neuronal response is poorly understood. Characterizing the stochasticity of gene expression is a prerequisite to understanding the fidelity and specificity of the process of translating environmental signals into a cellular response. A first step to comprehend how neurons control and/or benefit from the biochemical noise is to evaluate the cell-to-cell variability after a homogeneous stimulus.

2.2.2 Molecular mechanisms regulating activity-dependent transcription of immediate-early genes

As mentioned above, multiple activity-regulated transcriptional mechanisms coexist in the promoters of the IEGs, allowing a fast and robust stimulus-dependent induction in neurons. Fos was the first of these genes to be identified (Cochran et al.,
1984), and many studies have focused on the molecular mechanisms associated with this gene. In general, four classes of mechanisms have been observed:

- Transcriptional factors (TFs) pre-bound to stimulus-response elements in the proximal promoter.

- Stimulus-induced recruitment of co-activators and co-repressors of these TFs, which make post-translational modifications to the proximal histones (e.g. acetylation) and TFs (e.g. phosphorylation).

- Regulation by distant enhancer elements (EEs).

- Regulation of transcriptional elongation (i.e. RNA polymerase II pausing).

Additionally, the regulation of nuclear import (e.g. cytoplasmic sequestration of co-activators, and local tethering of TFs near translocation channels) also plays a role in the response regulation. The specific transcription factors and molecular components involved, as well as the conditions where these take part, are described in detail in Lyons and West (2011) and West and Greenberg (2011). A simple diagram exemplifying these mechanisms in the promoter before and after the stimulus is shown in Figure 2.5. In the case of sensory experience in neurons, the neurotransmitter triggers an increase in cytoplasmic calcium, which is known to activate the expression of IEGs (Greenberg et al., 1986). The inactivation mechanisms of the activity-dependent IEGs remain poorly understood, but recently some studies have started to explore these (Yang et al., 2016). Our analysis will focus on the first 30 minutes after gene expression induction, such that only the mechanisms of activation are relevant for the current work.
Figure 2.5: **Mechanisms regulating activity-dependent early gene transcription in neurons.** A simple diagram of the diverse mechanisms involved in the regulation of activity-dependent immediate-early gene transcription. Before neural activity, the promoter is primed for transcription with transcription factors (TF<sub>1,2</sub>) pre-bound to stimulus-response elements (gray boxes) and distal enhancer elements (EE), and the RNA-polymerase II (Pol II) paused in the promoter; additionally, some enzymes (E<sub>1</sub>) help to keep the repressed promoter state (e.g. histone deacetylase). Some leaky transcription (thin black arrows) might occur—which is expected in both directions—but the promoter is considered inactive (blunt red arrow). After neural activity—an increase in intracellular calcium ([Ca<sup>2+</sup>])—, some enzymes are released (E<sub>1</sub>), and others recruited (E<sub>2</sub>), making posttranslational modifications to both transcription factors (e.g. phosphorylation, P) and histones (e.g. acetylation, A), which induce gene transcription (thick black arrow). Long-distance looping of DNA and the proximity of associated enhancers (EE) have also been proposed as a mechanism of regulation of this activity-dependent transcription; in particular, eRNAs—short non-coding RNAs transcribed from enhancer regions—have been detected under neural activity. Now, the promoter is considered active (green arrow).

### 2.2.3 Single-molecule activity-dependent gene expression temporal dynamics in neurons

Our collaborators in the laboratory of Dr. Anne West (Neurobiology, Duke University) have used a highly-sensitive single-molecule RNA fluorescence in situ method (smFISH) to quantify the number of transcripts (mRNA) and active transcription sites (TSs) in individual neurons before and after induction of activity-dependent Fos gene expression. The cells used correspond to dissociated cultures of primary neurons from embryonic day 16 (E16) mouse hippocampus that have been grown in vitro for 7 days (DIV 7). Increasing extracellular potassium (KCl) in an isotonic solution was used to induce activity-regulated gene transcription. This induction
represents a well-established model to drive membrane depolarization, opening the L-type voltage-gated calcium channels, increasing the intracellular calcium ([Ca$^{2+}$]), which activates Fos and other IEGs.

The mRNA expression was quantified in cells at four time points: (1) just before adding the stimulus (uninduced conditions), (2) after five minutes of membrane depolarization (KCl (5 min) conditions), and (3) after 10 and (4) 20 minutes with the cells back in common condition medium (KCl (5 min) + CM (10 min) and KCl (5 min) + CM (20 min) conditions). In all cases, the cell samples were fixed and multiple (30-50) short (20bp) oligonucleotide probes complementary to the sequence of interest collectively bind individual mRNA molecules, allowing the automated identification and counting of absolute mRNA numbers in single cells using the image analysis program FISH-quant (Mueller et al., 2013). Additionally, the active transcription sites are detected because nascent mRNAs are transiently attached to the elongating RNA Polymerase II in the gene, accumulating fluorescent probes around active sites, and then appear as highly intense dots (1 or 2, as there are two copies of the gene) in the nucleus of the diploid cell (Figure 2.6). This was confirmed by using a second set of smFISH probes specific for the gene introns, which are only present in nascent mRNAs in the nucleus and are removed in mature mRNA molecules in the cytoplasm. These intron probes perfectly co-localized with the predicted transcription sites by the standard probes in dual-color smFISH experiments.

The experimental method was validated using multiple tests. First, the observed population average numbers of mRNA molecules matched the value quantified by quantitative real-time PCR, a well-established technique to quantify mRNA transcription levels in a population. Second, the fluorescence amplitude for individual dots detected in the smFISH experiments followed a skewed normal distribution, as observed in other smFISH experiments (Zenklusen et al., 2008; Mueller et al., 2013). Third, all cultured cells are expected to sense a homogeneous stimulus. This was con-
Figure 2.6: **Single-molecule mRNA fluorescence in situ method (smFISH).** Diagram of the smFISH method. Fluorescent labeled probes bind specifically to the gene of interest, either in the intron regions – detecting only nascent mRNAs – or in the exon regions – detecting both nascent and mature mRNAs. The nascent mRNAs accumulate in active promoters, appearing as intense dots in the nucleus – one for each active promoter; if two-colored labeled probes are used – one for the introns, other for exons – the resulting intense dots are expected to co-localize. Individual mature mRNA molecules appear as small fluorescent dots in the cytoplasm.

Confirmed by measuring intracellular calcium relative concentration in uninduced and induced conditions; in both cases, a smooth unimodal distribution was detected. This observation suggests that any cell-to-cell variability in gene expression might be arising from intrinsic stochasticity, and not from the existence of “micro-environments” or heterogeneous cell stimulation in the culture.

The experimental data to be analyzed here consists of 173 cells in *uninduced* condition, 174 cells in *KCl (5 min)* condition, 122 cells in *KCl (5 min) + CM (10 min)*, and 115 cells in *KCl (5 min) + CM (20 min)*. For each individual cell (1) the free mRNA molecule number, and (2) the number of active transcriptional sites (TSs) are known (Figure 2.7). Increasing the sample size is time consuming and expensive, because primary neurons are used for these experiments. Moreover, the time required for the set-up of the experiments prevents time measurements with less than a 5-minute interval. The model developed in the next section can extract meaningful kinetic and regulatory information from samples of approximately 150
Figure 2.7: Single-molecule fluorescent in situ hybridization (smFISH) measurements of Fos mRNA in neurons. Measurements are shown before the stimulus (uninduced), after 5 minutes in KCl stimulus, and additional 10 minutes and 20 minutes in common condition medium. (A) An example of an image-processed cell for each condition is shown, highlighting the detected mRNAs (green circles) and active transcription sites (TSs; green squares). The contour of the considered cell is marked (yellow line), as well as its nucleus (dashed yellow line). (B) For each condition, the histogram of number of free mRNA molecules detected per cell with a particular number of active transcription sites (TSs) are shown (blue dots, left y-axis), as well as a smoothed histogram with bins ±10 mRNA number (gray dots, right y-axis). The total number of cells per time sample is listed at the top (sample size, n).

cells at 4 time-point measurements.

2.3 Model of stochastic dynamics of early transcriptional response

2.3.1 Gene-expression probability distribution and its dynamics

The Chemical Master Equation (CME), also known as the Forward Kolmogorov equation, can be used to describe the system dynamics taking in account the discreteness of molecules and the stochasticity of biochemical events. Generally, the
CME is a linear differential equation describing the dynamics of the probability of
the system being in a particular state $x$ at time $t$, $P(x,t)$ (Zeron and Santillán, 2011).
More specifically, the CME is:

$$\frac{\partial P(x,t)}{\partial t} = \sum_k [a_k(x - \nu_k) \cdot P(x - \nu_k,t) - a_k(x) \cdot P(x,t)].$$

(2.1)

Here, the summation goes over all possible reactions, and each reaction $k$ updates
the system state as follows:

$$x \xrightarrow{a_k(x)} x + \nu_k$$

(2.2)

where $a_k(x) \, \partial t$ is the probability that the reaction $k$ will occur within the following
infinitesimal time interval $\partial t$, given that the system is at state $x$, and $\nu_k$ is the
stoichiometric vector of reaction $k$, i.e. how the system state changes when the
reaction $k$ occurs. Notice that for all $\{x,t\}$, $P(x,t) \geq 0$ and $\sum_x P(x,t) = 1$, given
that these are probabilities. In order to consider the time evolution of the probability
distribution for the whole system, or the ensemble of all possible states, this equation
can be written in matrix form:

$$\frac{\partial \mathbf{P}(\mathbf{X},t)}{\partial t} = \mathbf{A} \cdot \mathbf{P}(\mathbf{X},t),$$

(2.3)

where $\mathbf{X} = [x_1, x_2, \ldots]^T$ corresponds to a particular enumeration of all possible states
$x_i$, and $\mathbf{A}$ is a linear transformation called state reaction matrix with elements:

$$A_{ij} = \begin{cases} 
-\sum_k a_k(x_i) & \forall i = j \\
a_k(x_i) & \forall j \text{ such that } x_j = x_i + \nu_k \\
0 & \text{otherwise}
\end{cases}$$

(2.4)

Notice that this state reaction matrix has the following properties: (1) it is independent
of time, (2) all of its diagonal elements are non-positive ($A_{ii} \leq 0 \ \forall \ i$), (3) all
of its off-diagonal are non-negative ($A_{ij} \geq 0 \ \forall \ i \neq j$), and (4) all of its columns sum
exactly zero ($\sum_i A_{ij} = 0 \ \forall \ j$).
The CME framework can be used to describe (1) the cell-to-cell variability expected in the population, and (2) the population dynamics after stimulus under the described 2-states model (Figure 2.4). Assuming the population has reached an equilibrium state in the \textit{uninduced} condition, the stationary solution of the system CME (i.e. \( \partial P(X,t)/\partial t = 0 \)) represents the expected gene expression distribution in any sample. The stimulus is expected to disturb this equilibrium state because it changes the biophysical parameters of the model, and the system will evolve to a new stationary state. The time evolution of the probability distribution \( P(X,t) \) can be calculated using numerical methods to solve Equation 2.3 given an initial \( P(X,0) \) in the \textit{uninduced} condition.

In our system, the individual cell state \( (x) \) is defined by the number of active promoters (i.e. active transcription sites) \( [\rho_{ON}] \) and the number of mRNA molecules \( [m] \) present:

\[
x = \begin{pmatrix}
[\rho_{ON}] \\
[m]
\end{pmatrix}
\] (2.5)

where \( [\rho_{ON}] \in \{0, 1, 2\} \) and \( [m] \in \{0, 1, 2, 3, \ldots\} \); by definition \( [\rho_{OFF}] = 2 - [\rho_{ON}] \).

There are four possible reactions:

- **Promoter activation**: \( [\rho_{ON}] \rightarrow [\rho_{ON}] + 1 \)
  
  \[
  a_k(x) = k_{ON} \cdot (2 - [\rho_{ON}]) \\
  \nu_k = [1, 0]^T
  \]

- **Promoter inactivation**: \( [\rho_{ON}] \rightarrow [\rho_{ON}] - 1 \)
  
  \[
  a_k(x) = k_{OFF} \cdot [\rho_{ON}] \\
  \nu_k = [-1, 0]^T
  \]

- **mRNA synthesis**: \( [m] \rightarrow [m] + 1 \)
  
  \[
  a_k(x) = \mu_0 \cdot (2 - [\rho_{ON}]) + \mu \cdot [\rho_{ON}] \\
  \nu_k = [0, 1]^T
  \]

- **mRNA degradation**: \( [m] \rightarrow [m] - 1 \)
  
  \[
  a_k(x) = \delta \cdot [m] \\
  \nu_k = [0, -1]^T
  \]

A chemical reaction system is \textit{irreducible} if and only if for every pair of possible states \( x_i, x_j \) exists a finite set of reactions with nonzero probability such that it is possible for the system to go from state \( x_i \) to state \( x_j \). In other words, each possible state is reachable from any other possible state in the system. Zeron and Santillán (2011) have shown that any irreducible chemical reaction system has a unique and
Figure 2.8: Possible reactions and system state space. The cell’s state $x$ change over time is equivalent to a random walk on the lattice depicted above. The individual cell state $x$ is defined as the number of active promoters $\rho_{ON} \in \{0, 1, 2\}$ and the number of mRNA molecules $m \in \{0, 1, 2, \ldots\}$. There are four possible reactions: (1) Promoter activation (blue arrow), which increases $\rho_{ON}$ by one; (2) promoter inactivation (purple arrow), which decreases $\rho_{ON}$ by one; (3) mRNA synthesis (orange arrow), which increases $m$ by one; and (3) mRNA degradation (gray arrow), which decreases $m$ by one.

Asymptotically stable stationary distribution $\mathbf{P}^*$ such that

$$\frac{\partial \mathbf{P}^*(\mathbf{X}, t)}{\partial t} = \mathbf{A} \cdot \mathbf{P}^*(\mathbf{X}, t) = 0,$$

and corresponds to the nonzero eigenvector $\mathbf{V} \geq 0$ in the kernel of $\mathbf{A}$ after renormalization:

$$P^*(x_i) = \frac{V_i}{\sum_j V_j} \geq 0 \quad (2.7)$$

where $V_i$ is the $i$th element in the vector $\mathbf{V} = [V_1, V_2, \ldots]^T$, keeping $\sum_i P^*(x_i, t) = 1$. The reactions in our system allow moving individually in each dimension of the system state by $\pm 1$ (Figure 2.8), and then each system state is reachable from any other system state through a series of reactions, i.e. this system is irreducible. Then, a unique stationary probability distribution exists and can be calculated.

For cells in the uninduced state, the gene expression system was assumed to be in a stationary state, thus $\mathbf{P}^*$ is a good descriptor of the system initial distribution ($\mathbf{P}(\mathbf{X}, 0)$). I used numerical methods to calculate the stationary probability distribu-
bution ($P^*$); this requires the definition of a finite number of possible states, making $P$ and $A$ also of finite size. To do this, I assumed an upper limit for the values of $[m] \in \{0, 1, 2, ..., m_M\}$, such that $P(x, t) = 0 \forall t$ and $x$ with $[m] > m_M$. Then, $P^*$ is estimated by finding the eigenvector $V$ with eigenvalue $\lambda = 0$ for the state reaction matrix $A$ for any set of kinetic parameters $\theta$ using the Arnoldi iteration numerical method (Lehoucq and Sorensen, 1996). This method requires our matrix $A$ to be square and not symmetric, which is always true for our defined system.

The observed expression distributions change considerably in the first minutes after stimulus (Figure 2.7), which suggests that the system has not reached its new stationary state. Given the initial distribution $P(X, 0)$, we estimate the probability distribution dynamics $P(X, t)$ after induction using the Finite State Projection (FSP) method (Munsky and Khammash, 2006).

The FSP method allows the estimation of a probability distribution $P(X, t)$ at any specific time $t_f$ having an initial distribution $P(X, 0)$ and state reaction matrix $A$:

$$P(X, t_f) = \exp(A \cdot t_f) \cdot P(X, 0).$$  \hspace{1cm} (2.8)

If only a finite number of possible states exists (i.e. $P$ and $A$ have a finite size), then FSP gives an exact analytical solution. For infinite or extremely large number of states, FSP provides an approximate solution with guaranteed lower and upper bounds on the true solution (Munsky and Khammash, 2006). We used this method because it can be easily implemented to estimate the gene expression distribution dynamics after induction, while evaluating the approximation error as a function of $m_M$.

We define a specific model for the activity-dependent gene expression system by which biophysical parameters ($k_{ON}$, $k_{OFF}$, and/or $\mu$) are affected by the stimulus. The mRNA degradation rate $\delta$ and model structure (i.e. possible reactions and num-
ber of promoter states) stay constant for all cases. Then, for each specific model, a set of biophysical parameters \( \theta = \{ k_{ON}, k_{OFF}, \mu_0, \mu \} \) describing the system dynamics in uninduced (\( \theta_U \)) and after stimulus (\( \theta_S \)) conditions can be assigned. As described above, given the parameter values, we can determine the uninduced stationary probability distribution \( P_U(X, 0) \), with \( A(\theta_U) \cdot P_U(X, 0) = 0 \), and the probability distributions at each time measurement \( \tau \) after stimulus \( P_S(X, \tau) = \exp(A(\theta_S) \cdot \tau) \cdot P_U(X, 0) \), where \( A(\theta) \) represents the state reaction matrix for the particular set of parameters \( \theta \) (Figure 2.9). Our goal is to infer the parameter values \( \{ \theta_U, \theta_S \} \) that best describe the experimental data \( Y = \{ y_1, y_2, \ldots \} \), i.e. a set of observations (individual cells) \( y_i = \left( \begin{array}{c} [P_{ON}] \\ m \end{array} \right) \in X \).

2.3.2 Estimating kinetic parameters: A Bayesian approach

In order to determine the model and biophysical parameters that best describe the biological phenomenon of interest, I have implemented a Bayesian approach with two aims:

1. **Parameter inference**: Use the posterior probability distribution \( P(\theta|Y, N) \) of the biophysical parameter sets \( \theta = \{ \theta_U, \theta_S \} \) given the observed experimental data \( Y \) and the model \( N \) to infer these parameter values \( \{ \theta_U, \theta_S \} \) that best explain the data, and evaluate the confidence on these values.

2. **Model comparison**: If there are multiple 2-states models, then infer which model \( N \) is most plausible given the observed experimental data \( Y \), penalizing by the complexity of the model.

Bayesian methods, in contrast to standard optimization methods such as maximum likelihood, naturally quantify the uncertainty in the estimated parameters. The essential characteristic of Bayesian methods is their explicit use of probability
Figure 2.9: Example of a specific model for the activity-dependent gene expression system. An example with only $k_{ON}$ sensitive to the stimulus, and biophysical parameter values: $k_{OFF} = 0.0459 \text{ min}^{-1}$, $\mu_0 = 0.026 \text{ mRNA/min}$, $\mu = 5.65 \text{ mRNA/min}$, $k_{ON}^{(U)} = 0.0049 \text{ min}^{-1}$, $k_{ON}^{(S)} = 0.1110 \text{ min}^{-1}$. For the uninduced condition and each time point measurement after stimulus the probability distributions (pink lines), and an example of a random sample $Y$ of 150 cells per condition (blue dots), are shown.

distributions $P(\{\theta_U, \theta_S\})$ of all parameter “solutions” that are in agreement with the observation, instead of searching for a single “best” solution $\{\hat{\theta}_U, \hat{\theta}_S\}$ (Gelman et al., 2013). Having a probability distribution associated with the biophysical parameter values describing the observed data allow us to utilize statistical analysis on it, directly evaluating parameter identifiability and the confidence on the model. In general, to apply a Bayesian approach, we need to determine a full probability model to calculate the associated posterior probability distribution, and to evaluate the fit of the model.
The posterior probability is defined as the probability of the parameter set $\theta$ given some evidence—observed data—$Y$ and assuming a specific model $N$:

$$P(\theta|Y,N) = \frac{P(Y|\theta,N) \cdot P(\theta|N)}{P(Y|N)}.$$ (2.9)

$P(Y|\theta,N)$ is known as the likelihood of the data $Y$ given the parameters and model; $P(\theta|N)$ represents the prior belief on the parameter set $\theta$; and $P(Y|N)$ is known as the evidence of the data $Y$ under the particular model $N$. Notice that this evidence is the same for all parameter sets $\{\theta_U, \theta_S\}$, and consequently it does not affect the shape of the posterior distribution. Analogously, if no discriminatory information is available, all parameter sets can be assumed equally probable, making the prior $P(\theta|N)$ a constant value. Then, in this case, simulating the likelihood distribution is equivalent to simulating the posterior distribution.

In the context of our model, the experimental data is composed by a sample of $n_U$ cells in the uninduced condition, and $n_\tau$ cells at each following time point measurement $\tau \in T$ after induction, and the state of each individual cell is $y \in X$. Then, the likelihood of observing the complete data sample $Y$ follows a Multinomial distribution:

$$L(Y) = P(Y|\theta,N) = \frac{n_U!}{\prod_k n_{U,k}!} \prod_{i=1}^{n_U} P_{U}(y_i|\theta_U,N) \cdot \prod_{\tau \in T} \frac{n_\tau!}{\prod_k n_{\tau,k}!} \prod_{i=1}^{n_\tau} P_{\tau}(y_i|\theta_S,N)$$ (2.10)

where $n_{*,k}$ is the number of cells in the sample having the state $k \in X$ ($\sum_k n_{*,k} = n_*$), $P_U = P_U(X,0)$ is the stationary distribution associated with the uninduced state, $P_\tau = P_S(X,\tau)$ is the probability distribution $\tau$ minutes after induction, and $T$ is the set of time points measured. Here, we are assuming the cells in the sample are independent and identically distributed (iid). For simplicity, the logarithm of the
likelihood is generally used:

\[
\ln \mathcal{L}(Y) = \ln P(Y|\theta, N) = \ln(\eta_U) \sum_{i=1}^{n_U} \ln(P_U(y_i|\theta_U, N)) + \sum_{\tau \in T} \ln(\eta_\tau) \sum_{i=1}^{n_\tau} \ln(P_\tau(y_i|\theta_\tau, N)),
\]

where \( \eta_\tau = n_\tau! / \prod_k n_{\tau,k}! \) is the multinomial coefficient for each sample.

In general, the posterior probability distribution cannot be solved analytically, and computational methods must be applied (Liepe et al., 2014). Markov-chain Monte Carlo (MCMC) methods allow a systematic sampling of any probability distribution, extracting the desired information. In general, these methods are based on the principle that if we can draw a finite number of independent and identically distributed (\textit{iid}) samples from the corresponding posterior distribution, then the properties of this finite sample will approximate the properties of the posterior. To generate these \textit{iid} samples, one of the simplest and more used methods is the Metropolis Random Walk (MRW) algorithm (Metropolis et al., 1953; Hastings, 1970), which in general terms is:

1. Choose some initial parameters \( \theta^{(1)} \) and calculate the corresponding likelihood \( \mathcal{L}^{(1)} = P(Y|\theta^{(1)}, N) \).

2. Iterate over \( t = \{1, 2, ..., t_{MAX}\} \) as follows:

   (a) Draw a random proposal \( \phi \sim \theta + \mathcal{N}(|\theta|)(0, \Sigma) \), where \( \mathcal{N}(|\theta|)(0, \Sigma) \) is a Multivariate Normal distribution with the same dimension as \( \theta \), mean zero and covariance matrix \( \Sigma \).

   (b) Calculate the likelihood of the proposal \( \mathcal{L}_\phi = P(Y|\phi, N) \). If any parameter value is outside the model support (i.e. negative numbers), the probability cannot be explicitly calculated, then assign \( \mathcal{L}_\phi = 0 \).
Figure 2.10: Example of a run of the Metropolis Random Walk algorithm in our system. (A) The log-likelihood per iteration, highlighting the “burn-in” period (red dotted line). (B) The selected parameter values per iteration after the “burn-in” period. In this example, the 2-states $k_{ON}$-sensitive model is used (see Appendix A for implementation details).

(c) Update parameters $\theta_{(t+1)} \leftarrow \phi$ with probability $\min(1, \frac{L^*}{L_{(t)}})$; otherwise, $\theta_{(t+1)} \leftarrow \theta_{(t)}$.

Notice that because the algorithm evaluates the ratios, it is independent of normalizing constants, as the evidence factor in the posterior probability (Equation 2.9) and the multinomial coefficients $\eta$ in the likelihood function (Equation 2.10). As the Markov chain evolves, the likelihood will improve and eventually plateaus. The initial period is termed the “burn-in” period, and these samples are discarded from the final statistics. An example of a run of this algorithm in our system is shown in Figure 2.10. See Appendix A for details on the implementation and code.
2.3.3 Comparing models: Information Criterion metrics

Ideally, we would like to infer which model is the most plausible given the data:

\[ P(N|Y) \propto P(Y|N) \cdot P(N) \tag{2.12} \]

where \( P(N) \) represents the prior belief for the model, and the evidence corresponds to:

\[ P(Y|N) = \int P(Y|\theta, N) \cdot P(\theta|N) \cdot d\theta, \tag{2.13} \]

where the term \( P(\theta|N) \) is known as the Occam factor. This factor can be understood as a measure of the complexity of the model, depending not only on the number of parameters in the model, but also on the prior probability that the model assigns to them. This term also relates to the amount of information we gain about the model’s parameters when the data arrive (MacKay, 2003). The model which achieves the greatest evidence is a trade-off between minimizing this natural complexity measure –Occam factor– and maximizing the data’s likelihood. Calculating this integral directly is usually impossible analytically, and very expensive computationally.

More generally, one can compare models by contrasting the model fit and the model complexity. One option is applying simple likelihood-based metrics that directly penalize the number of free parameters, as:

- **Bayesian Information Criterion:**

  \[ BIC = -2 \cdot \ln(\hat{L}) + k \cdot \ln(n), \tag{2.14} \]

  where \( \hat{L} \) is the maximized likelihood of the model, \( k \) is the number of free parameters, and \( n \) is the total sample size, assuming \( n \gg k \). The model with the lowest BIC value is preferred (Schwarz, 1978).

- **Akaike Information Criterion:**

  \[ AIC = 2 \cdot k - 2 \cdot \ln(\hat{L}), \tag{2.15} \]
where $\hat{L}$ is the maximized likelihood of the model, and $k$ is the number of free parameters. It is also possible to correct by sample size

$$AIC_c = AIC - \frac{2k(k+1)}{n-k-1},$$

(2.16)

where $n$ is the total sample size. Also the model with the lowest AIC value is considered to represent better the process that generates the data (Akaike, 1998).

Another alternative is to take advantage of the performed MCMC analysis, with no approximations necessary, using the posterior distribution—a "Bayesian measure"—to evaluate both the fit of the model and its complexity:

- Deviance Information Criterion:

$$DIC = p_D + D,$$

(2.17)

where $p_D$ represents the effective number of parameters of the model,

$$p_D = E_{\theta|Y}[D(\theta)] - D(E_{\theta|Y}\theta)$$

(2.18)

with $E_{\theta|Y}$ representing the expected value over the posterior probability distribution, and $D(\theta)$ is known as the deviance and represents the goodness of fit,

$$D(\theta) = -2 \cdot \ln(P(Y|\theta)).$$

(2.19)

This metric is intended as a generalization of AIC, and is only valid when the posterior is approximately multivariate normal. The minimum DIC estimates the model that will make the best short-term predictions (Spiegelhalter et al., 2002).

The three metrics will be used to compare the models in our system (see Section 2.4.2).
2.4 Characterizing the kinetic parameters of *Fos* transcriptional response in neurons

The activity-dependent IEGs promoter is assumed to be in one of two states (2-states model): (1) an *inactive* state (*ρ*\(_{\text{OFF}}\)), which dominates in the condition without neural activity or stimulus, where Pol II is paused at the promoter, and only a small –leaky– transcription level (synthesis rate *µ*\(_0\)); and (2) an *active* state (*ρ*\(_{\text{ON}}\)), which dominates in the condition with neural activity or stimulus, where all the necessary molecular machinery is bound to the gene promoter with the proper chromatin structure, which results in Pol II elongation and a high transcription level (synthesis rate *µ* ≫ *µ*\(_0\)). In reality, many small steps or molecular events must take part to switch the promoter from inactive to active state (Figure 2.5); but the order and temporal scale of these molecular processes are currently unknown. We expect that at least one limiting –slow– step exists in the activation process because after the stimulus cells often display just one allele active. On the other hand, given the *primed* state of the promoter and the observed fast activation (see Section 2.2), most of these events are expected to occur rapidly, potentially behaving as a 2-states model.

As mentioned in Section 2.1, the transitions between promoter states are assumed to occur stochastically and independently in each promoter. This assumption is supported by the detection of many cells with only one active promoter (one detected TS) in the experimental measurements (Figure 2.7B). Also, postmitotic neurons used for the experimental measurements do not enter cell cycle, and two copies of the promoter per diploid cell are considered at any given time ([*ρ*\(_{\text{ON}}\)] + [*ρ*\(_{\text{OFF}}\)] = 2 per neuron). The previously measured mRNA degradation rate of the *Fos* gene was used, *δ* = 0.0462 min\(^{-1}\) (Shyu et al., 1989).

The stimulus increases the activity-dependent genes expression level in neurons (Greenberg et al., 1986) (Figure 2.7), but the exact molecular mechanism responsible
for this increment on mRNA molecules is currently unknown. The effect of these different scenarios in the system, and which one represents better the experimental data, is discussed in Section 2.4.

2.4.1 2-states model with sensitive activation rate

Initially we considered the model where only the promoter activation rate $k_{ON}$ is sensitive to stimulus, i.e. all the biophysical parameters are exactly the same in the uninduced and after stimulus conditions, except for $k_{ON}$. We chose this model first because the polymerase is expected to be paused in the promoter and released after stimulus, given the canonical model of the activity-dependent genes (Figure 2.5). This can be interpreted as an increase on the activation rate. I ran three replicas with different random initial conditions of the MRW algorithm for 100,000 iterations (the range of values considered for the initial conditions and other algorithm details are specified in Appendix A).

All replicas converged to practically identical biophysical parameter distributions (Figure 2.11A), suggesting a unique optimal solution exists. Moreover, the joint distribution of individual biophysical parameters and the log-likelihood associated with each parameter set ($\ln L_\theta$) shows a well-defined peak in all cases (Figure 2.11C), suggesting once again that an optimal value exists. In the deterministic version of this model, the steady states values depend on the ratio, instead of the individual values, of the activation and deactivation rates ($\bar{m} = \frac{k_{ON}}{k_{ON} + k_{OFF}} \frac{\mu}{\delta} + \frac{k_{OFF}}{k_{ON} + k_{OFF}} \frac{\mu_0}{\delta}$), and the individual $\{k_{ON}, k_{OFF}\}$ values could not be deduced from $\bar{m}$. Our method is capable of efficiently identify the values of these parameters because, as mentioned above, different transition rates ($\{k_{ON}, k_{OFF}\}$) can result in qualitatively different expression distributions even if the average mRNA levels $\bar{m}$ do not change (Figures 2.2 and 2.4). Additionally, the probability distribution dynamics ($P_S$) are particularly sensitive to the magnitude of these parameters, relaxing to the stationary state faster as the
numbers increase.

The resulting unimodal posterior distributions and the consistency among replicas supports the robustness of the described approach. However, it does not show anything about the reliability of the specific parameter values obtained. As mentioned in the previous section (2.3.2), the main advantage of the Bayesian approach is that we can evaluate the posterior distributions. All the biophysical parameters resulted in narrow distributions with small fold change variation, with the exception of some outliers (Figure 2.12). The distribution was considerably more narrow for the synthesis rate $\mu$ than the parameters describing the bursts dynamics (i.e. $\{k_{ON}, k_{OFF}\}$). The gene expression distributions were calculated for all 2,807 unique parameter sets $\theta$ obtained from all three replicas, weighting by the occurrence of each $\theta$ (Figure 2.13). No significant variation is observed for neither $P_U$ or the multiple $P_S$ distributions, and all $\theta$’s show the same qualitative behavior (see Appendix A for implementation details).

2.4.2 Other models: changing the sensitive parameters

I applied the same method to other 2-states models by varying the parameters that could be affected by the stimulus:

- $k_{OFF}$-sensitive,
- $\mu$-sensitive,
- $k_{ON}, k_{OFF}$-sensitive,
- $k_{ON}, \mu$-sensitive,
- $k_{OFF}, \mu$-sensitive,
- $k_{ON}, k_{OFF}, \mu$-sensitive.
Figure 2.11: Biophysical parameter inference under a 2-states $k_{ON}$-sensitive model. (A) Marginal posterior distributions for the biophysical parameter values for three MRW replicas; $k_{ON}^{(U)}$ refers to the $k_{ON}$ value in uninduced conditions ($\theta_U$) and $k_{ON}^{(S)}$ refers to the $k_{ON}$ value in after stimulus conditions ($\theta_S$). (B) Joint distributions of pairs of parameter values. (C) Joint distribution of parameter values and log-likelihood. (B-c) use the results from all replicas.
Figure 2.12: Variance in parameter inference under a 2-states $k_{ON}$-sensitive model. Fold change variation and coefficient of variation (i.e., variance over mean) observed in the individual parameters in the posterior distributions from Figure 2.11. In each boxplot, the central mark is the median, the edges of the box are the 25th and 75th percentiles, the whiskers extend to the most extreme data points not considered outliers, and outliers are plotted individually.

The resulting parameter distributions were compared (Figure 2.14), and the models were evaluated with different Information Criteria metrics (see Section 2.3.3).

The three metrics used—Bayesian Information Criterion (BIC), Akaike Information Criterion (AIC), and Deviance Information Criterion (DIC)—showed exactly the same qualitative behavior (Figure 2.15). Interestingly, the lowest scores occur when the promoter activation rate ($k_{ON}$) is affected by the stimulus. The best model was the $k_{ON}, k_{OFF}, \mu$-sensitive model. Nevertheless, the MRW showed a bimodal distribution on the values of $\mu_0$, and a trend to higher synthesis rate in the inactive promoter $\mu_0$ than in the active promoter $\mu$, which does not make sense biologically (Figure 2.16). The $k_{ON}, k_{OFF}$-sensitive model showed a good score, but only slightly better than the $k_{ON}$-sensitive model (Figure 2.15 and Figure 2.17). The $k_{ON}$-sensitive model is the strongest and most parsimonious model representing the experimental data.
Figure 2.13: Probability distribution associated with parameter estimates under a 2-states \(k_{\text{ON}}\)-sensitive model. Resulting variation in the corresponding gene expression probability distributions for the \(\theta\)'s obtained in the MRW from Figure 2.11. The experimental data \(Y = \{y_i\}\) is shown for comparison (gray dots).

2.5 Using synthetic data to validate estimation and propose a better sampling approach

If we could measure 1000 cells and/or 6 time points, how much would the confidence on our parameter estimates improve? We can analyze this repeating the parameter fitting process on random samples generated from a known model, and testing different sample sizes and time point measurements. Knowing the biophysical parameter values used to generate the data, the deviation of the estimates can be directly calculated and compared. Then, an optimal sample strategy can be suggested for future experiments.

I used the \(k_{\text{ON}}\)-sensitive model and its best fit biophysical parameters \(\hat{\theta}\) to generate the synthetic data. The biophysical parameters with the highest posterior
Figure 2.14: Comparing parameter distributions of different 2-states models. The marginal posterior distribution for each biophysical parameter. Each row corresponds to a different model with the parameter(s) sensitive to stimulus shown in the left. When the parameter is sensitive to stimulus, uninduced conditions are shown as continuous lines, after stimulus conditions as dashed lines. The results from three MRW replicas are used in all cases.

Probability ($\hat{\theta}$) in the MRW run of the 2-states $k_{ON}$-sensitive model were:

$$\hat{k}^{(U)}_{ON} = 0.0049 \text{ min}^{-1}$$

$$\hat{k}^{(S)}_{ON} = 0.1110 \text{ min}^{-1}$$

$$\hat{k}_{OFF} = 0.0459 \text{ min}^{-1}$$

$$\hat{\mu}_0 = 0.0260 \text{ mRNA/min}$$

$$\hat{\mu} = 5.65 \text{ mRNA/min}$$
Figure 2.15: **Comparing different 2-states models by Information Criteria.** Three distinct metrics are applied—Bayesian Information Criterion (BIC), Akaike Information Criterion (AIC), and Deviance Information Criterion (DIC)—to the results obtained from the MRW to 2-states model where only the parameter(s) listed in the x-axis are sensitive to the stimulus.

Three random samples were generated for each “strategy” (i.e. time points $T$, and sample size $\{n_U, \{n_r\}_T\}$) from the associated probability distributions $\{\hat{P}_U, \{\hat{P}_r\}\}$, and this was followed by three independent MRW runs for 35,000 iterations performed on each of them (all other algorithm details as in Section 2.4; see Appendix A).

First, the experimental sample was replicated, i.e. $T = \{0, 5, 15, 25\}$ min, and $\{n_U = 173, n_{05} = 174, n_{15} = 122, n_{25} = 115\}$ cells. Second, I explored the effect of simply increasing the number of cells per time point, with $n = 300 \times T$ and $n = 1000 \times T$ with the same $T = \{0, 5, 15, 25\}$ min. Third, I explored the effect of different temporal sampling, while keeping a similar total number of cells sampled as in the experimental measure: $T = \{0, 5, 10, 15, 20, 25\}$ min with $n = 100 \times T$, $T = \{0, 5, 15\}$ min with $n = 200 \times T$, $T = \{0, 5, 25\}$ min with $n = 200 \times T$, and $T = \{0, 15, 25\}$ min with $n = 200 \times T$. A comparison of the posterior distributions obtained for each of these sampling strategies is shown in
Figure 2.16: **Biophysical parameter inference for $k_{ON}, k_{OFF}, \mu$-sensitive model.** Marginal and joint posterior distributions for the log-likelihood and biophysical parameter values for three MRW replicas; $p^{(U)}$ refers to the parameter value in *uninduced* conditions ($\theta^U$) and $p^{(S)}$ refers to the parameter value in *after stimulus* conditions ($\theta^S$) in each case.

Figure 2.18. The sample replicating the experiment shows distributions with a similar width than the experiment sample (Figure 2.11A), but variation in the modes appear from sample to sample.

In order to quantify the effect of the sample strategy, I calculated the mean
Figure 2.17: Biophysical parameter inference for the $k_{ON}, k_{OFF}$-sensitive model. Marginal and joint posterior distributions for the log-likelihood and biophysical parameter values for three MRW replicas; $k_{OFF}^{(U)}$ refers to the $k_{OFF}$ value in uninduced conditions ($\theta_U$) and $k_{OFF}^{(S)}$ refers to the $k_{OFF}$ value in after stimulus conditions ($\theta_S$).

development of the posterior distribution to the true biophysical parameter values $\hat{p} \in \hat{\theta}$:

$$
\langle \epsilon \rangle = \frac{1}{||MRW||} \sum_{p \in \theta} \sum_{i \in MRW} |p_i - \hat{p}|
$$

(2.20)

where $i$ is each iteration of the corresponding MRW run excluding the “burn-in” period, and $||MRW||$ is the total number of iterations considered (Figure 2.19).
Figure 2.18: **Sample strategy effect on posterior probability distributions.** The marginal posterior probabilities per biophysical parameter per sample strategy are shown for three random samples. The results from three independent MRW runs—different random initial conditions—were mixed for each case, excluding their “burn-in” period. In all cases, the x- and y-axes have exactly the same limits in each column and row, respectively, and the true parameter value $\hat{p} \in \hat{\theta}$ are shown as a dashed vertical line.
Using this metric, the experimental sampling strategy displayed higher variation among samples; and increasing the sample size to more than double \((n = 300 \times T)\) showed improvement only on average, but not significantly different from the experimental sample size. Having \(n = 1000 \times T\) reduces the sample variation, as well as significantly decreasing the deviation from the true values of the posterior distributions. However, this sample size is practically impossible in neurons with the current technology. Alternatively, increasing the number of time points while keeping the same number of total sample cells reduces the variation among samples, and has a small improvement in \(\langle \epsilon \rangle\). Interestingly, decreasing the number of time points by using only the earlier and later measurements after induction has a similar effect, but the improvement is lost if other time points are chosen. A more systematic exploration is required for deeper conclusions.

### 2.6 Computational cost & model complexity

The implementation of the algorithm is currently limited by the time required to run the MRW algorithm. Each iteration of the MRW requires the calculation of a stationary probability distribution (solving Eq. 2.6) and an implementation of the Finite State Projection method (solving Eq. 2.8; see Section 2.3.2). The run time of both of these operations depends on the state reaction matrix \(A\) size \((a \times a)\). Depending on the number of possible promoter states \((p)\) and the limit for the maximum mRNA number used \((m_M)\), the matrix size is:

\[
a^2 = \left( \frac{p^2 + p}{2}(m_M + 1) \right)^2 \sim O(m_M^2p^4). \tag{2.21}
\]

For example, in our 2-states model with \(m_M = 300\), the number of possible states for the system is \(a = (3)(301) = 903\), and the matrix size is 815,409; for a 3-states model, \(a = (6)(301) = 1806\), and the matrix size scales to 3,261,636.
Figure 2.19: Mean deviation $\langle \epsilon \rangle$ per sample strategy. The mean deviation of the posterior distribution $\langle \epsilon \rangle$ per sample strategy (Eq. 2.20) for three random samples (gray circles) are shown; the results from three independent MRW runs -different random initial conditions- were mixed for each case, excluding their “burn-in” period. The average of the independent samples is also shown for each strategy (squares): replicating the experimental sample (red), increasing the sample size (orange), and 600 cells in the total sample varying the time point measurements (green).

The calculation of the stationary distribution requires the eigendecomposition of the state reaction matrix $\mathbf{A}$. The eigendecomposition time or computational complexity scales as $O(a^3)$. In a standard PC, using MATLAB, the calculation of the stationary distribution for our 2-states model takes approximately 0.2 seconds, and scales up to 1.2 seconds under the 3-states model.

The FSP calculation requires the calculation of the exponential of the state reaction matrix $\mathbf{A}$ and a matrix-vector multiplication for each time point evaluated after stimulus (Eq. 2.8). The time cost of each multiplication step scales with the matrix size as $O(a^2)$. The matrix exponential using the scaling and squaring approximation (Higham, 2005) requires a matrix equation solution ($\sim O(a^3)$) and
\[ \pi_m + \left\lfloor \log_2(||A||_1/\theta_m) \right\rfloor \] matrix multiplications (\( \sim O(a^3) \)), where \( \pi_m \) and \( \theta_m \) are approximation constants, and \( ||A||_1 = \max_{1 \leq i \leq a} \sum_{j=1}^a |A_{ij}| \) is the 1–norm of the \( A \) matrix. The matrix norm depends on the reactions considered, the actual biophysical parameter values, as well as the \( m_M \) limit value. For example, assuming that each promoter can only be activated or inactivated (even if more than two promoter states are available; i.e. \( \rho_1 \equiv \rho_2 \equiv \ldots \equiv \rho_p \)), the matrix norm is less or equal to:

\[
||A||_1 \leq 2 \cdot (2 \cdot \max(\kappa) + 2 \cdot \max(\mu_s) + \delta \cdot m_M),
\] (2.22)

where there is two alleles or promoter copies per cell, \( \kappa \) is the set of promoter activation and inactivation rates, and \( \mu_s \) is the set of synthesis rates per promoter state. In a standard example of our 2-states model, the run time in a PC of the matrix exponential is approximately 4.15 seconds, and the total calculation of time for the probability distribution dynamics is just slightly higher (approximately 4.16 seconds).

Then, a MRW with 100,000 iterations as implemented before took approximately:

\[
100000 \times (0.2 + 4.16) \text{ seconds} \approx 5 \text{ days}.
\]

If the matrix size changes to \( a_* \), the computational time is expected to be roughly:

\[
100000 \times (0.2 + 4.16 \cdot \frac{\alpha_*}{\alpha} \cdot \frac{a_*^3}{903^3}) \text{ seconds},
\]

where \( \alpha = \pi_m + \left\lfloor \log_2(||A||_1/\theta_m) \right\rfloor \) for the standard model used to calculate the running time per operation, and \( \alpha_* \) is the equivalent value for the new state reaction matrix. For example, if a 3-states model is used with the same \( m = 300 \), the expected running time is approximately:

\[
\frac{6^3}{3^3} \times 5 \text{ days} \approx 40 \text{ days},
\]

assuming the biophysical parameters describing the model have not change significantly (i.e. \( \frac{\alpha_*}{\alpha} \approx 1 \)). On the other hand, the same 2-states model but with a higher
maximum mRNA limit $m_M = 500$ would take:

$$\frac{501^3}{301^3} \times 5 \text{ days} \approx 23 \text{ days}.$$ 

2.7 Discussion & Conclusions

The biological question addressed in our collaboration with the laboratory of Dr. Anne West presented a series of particular challenges. As described in the introduction, the main reason to implement single molecule measurements of gene expression in neurons was to characterize the cell-to-cell variation associated with Fos and other activity-dependent immediate-early genes (IEGs) expression. I developed a model that takes into account the intrinsic stochasticity of the system resulting from the discreteness of molecules and the probabilistic nature of the biochemical events. The dynamics of induction and the population variability was determined by directly modeling the dynamics of the probability distribution using the Chemical Master Equation (CME). Finally, the small sample size of single-cell measurements made it impossible to directly fit biophysical parameter values to the sparse observed distributions. We overcame this problem by implementing a Bayesian approach, using the likelihood of the experimental data given the model to detect the most probable biophysical parameters. The Bayesian approach provides a direct measure of the confidence on the specific parameter values, and also allows for the evaluation of different model assumptions.

Using our algorithm, we can extract meaningful biological information from the small, noisy sample of Fos gene expression in single neurons (Figure 2.7). We showed that intrinsic stochasticity associated with a 2-states model (i.e. individual promoters can be either in active or inactive states) is enough to recapitulate the observed cell-to-cell variability (Figure 2.13). The model also predicts that the stimulus must regulate the promoter activation rate ($k_{ON}$) in order to explain the population dis-
tribution dynamics after stimulus. Moreover, the proposed model ($k_{ON}$-sensitive) exhibits a high parameter *identifiability*, i.e. the biophysical parameter values appear uniquely constrained by the data, giving confidence to the estimates. The confidence was further evaluated by simulating synthetic data from inferred parameters, and calculating the deviance of the parameter estimates from the “true” model. Our algorithm exhibited a robust behavior, converging quickly to the same posterior probability distribution starting from distinct random initial conditions.

An important assumption in our model is that the promoters have only two possible discrete states. The method framework can be easily extended to a 3-states model, in which each promoter state is described by a different synthesis rate and specific transition rates to the other states. The implementation of the algorithm is currently limited by the computational power, particularly by the calculation of the matrix exponential required to evaluate the gene expression distribution dynamics after stimulus (see Equation 2.8 and Section 2.6). A future possibility is to implement the algorithm in c++ instead of Matlab (which is currently used, see Appendix A), which is in general more stable, open source, and many instances can be simultaneously run in a computer cluster. I have already implemented the current version of the algorithm in c++, but the matrix exponential function resulted slower than Matlab implementation, which is known to be optimized for matrix operations. Nevertheless, if an *ad hoc* matrix exponential function is developed, the benefits of c++ must be revisited. Future work needs to explore these and other possibilities.

The stochastic dynamics of 2-states models of gene expression has been previously explored in diverse systems and conditions using distinct methods (see Appendix B). For example, Zenklusen et al. (2008) explored bursting dynamics in yeast by analyzing the mRNA expression distribution and its kinetic parameters of a set of canonical genes. The authors defined a mathematical model and used Monte Carlo simulations and analytical tools to calculate probability distributions associated with the pop-
ulation behavior of their smFISH gene expression data (e.g. frequency of active promoters). Another interesting study was performed by Senecal et al. (2014). The authors explored the effect of transcription factors (TFs) regulation over the early response gene Fos in human tissue using smFISH and immunofluorescence (IF) measurements. Senecal et al. (2014) observed that cells can use the TF concentration to tune the frequency of transcription bursts. They characterized the burst dynamics in this system by fitting biophysical parameters through several steps: first, fitting the distribution of active TSs to a binomial distribution, then using FSP algorithm to describe the nascent mRNA distribution, and finally choosing the values that better reproduce the observed mRNA distribution using Monte Carlo simulations (i.e. Gillespie algorithm). Recently, Sepulveda et al. (2016) used also smFISH and IF single-cell measurements to study the role of promoter configurations on mRNA expression in the bacteriophage lambda system in Escherichia coli cells. The authors used FSP algorithm and maximum likelihood fitting to characterize the dynamics and stochasticity arising from the multiple promoter states. In general, these studies took advantage of well-behaved systems, and/or large population samples to characterize the gene expression dynamics and cell-to-cell variability. Applying population parameters to a general model might not be the best strategy when the population distribution is sparse and highly variable, such as the data presented in this chapter (Figure 2.7). This is the first work to simultaneously fit the promoter regulation and gene expression level, including the population distribution (i.e. intrinsic variability), and apply a Bayesian approach to get a sense of the confidence on parameter estimates and compare model assumptions. This potentially gives a more faithful representation of the underlying dynamics, which can be applied to a variety of systems.
2.8 Contributions

Chapter 2 was a collaboration between Liang-Fu Chen from the West lab and Mariana Gomez-Schiavon from the Buchler lab. All experimental measurements were performed by Liang-Fu Chen. The model design and mathematical analysis were done by Mariana Gomez-Schiavon. This work was supported by a CONACYT graduate fellowship (MGS), the National Institutes of Health Directors New Innovator Award DP2 OD008654-01 (NEB), the Burroughs Wellcome Fund CASI Award BWF 1005769.01 (NEB), the National Institutes of Health Exploratory/Developmental Research Grant Award R21DA041878 (AEW), and seed funding from the Duke Center for Genomic & Computational Biology (AEW and NEB).
Modeling evolutionary dynamics of epigenetic switches in fluctuating environments

Populations need to adapt constantly to new environmental conditions in order to survive. Traditionally, adaptation has been thought to occur simply through the emergence and selection of random mutations. Nevertheless, recent evidence suggests that the adaptation process and mechanisms involved might be considerably more complicated (Rosenberg, 2001; Jablonka and Lamb, 2005). For example, cases of adaptive mutation, where genetic variation occurs in response to the environment (Perfeito et al., 2007), directed mutation, where the useful mutations are induced preferentially (Sniegowski and Lenski, 1995), as well as cryptic variation, where mutations are accumulated and maintained without phenotypic consequences (Rohner et al., 2013), have been observed in diverse organisms across the tree of life. Additionally, the relevance of epigenetic inheritance in the adaptation process has been shown both theoretically (Day and Bonduriansky, 2011) and experimentally (Halfmann et al., 2012; Balaban et al., 2004; Beaumont et al., 2009).

The adaptation process in fluctuating environments is an old question in evo-
olutionary biology (Kimura, 1954; Cohen, 1966; Gillespie, 1972). In particular, the phenomenon of adaptive variation, where a population confronts uncertain environments by generating phenotypic variation among individuals in a population (Meyers and Bull, 2002), has been largely studied from different perspectives. This adaptive variation could arise through DNA mutation (genetic adaptation) that permanently modifies the phenotype, or it could emerge via spontaneous transitions between different pre-existing phenotypic states (epigenetic adaptation).

Epigenetic adaptation was first recognized as an important source of population variation by Nanney (1958). The benefit of phenotypic variation or \textit{bet-hedging}, when a percentage of the population switches to an alternative state before an upcoming environmental change, has been established for a long time (Cohen, 1966; Slatkin, 1974; Bull, 1987). However, Jablonka et al. (1992) first proposed a possible selective advantage of epigenetic adaptation as a mechanism to deal with fluctuating environments. \textit{Epigenetic switches} are systems capable of sustaining two heritable gene expression states without underlying DNA mutation (Iliopoulos et al., 2009); transitions between states can occur spontaneously due to stochastic fluctuations. In this case, epigenetic switches can work as mechanisms of bet-hedging. As explained in the introduction (see Section 1.4.2), an epigenetic switch can emerge through a simple positive feedback loop in gene networks, which helps explain why epigenetic switches are ubiquitous across both eukaryotes and prokaryotes with functions ranging from development to stress response (Casadesús and Low, 2006). This level of commonality suggests that epigenetic switches can be selected during evolution. Nevertheless, the specific evolutionary conditions that lead to the selection of epigenetic switches remain to be elucidated.

Epigenetic switches can sustain two states of gene expression and spontaneous transitions between states –bet-hedging– can allow for quick adaptation upon environmental change if one of the expression states is favored in the new environment.
Experiments have shown that stochastic phenotypic switching occurs in natural populations (Balaban et al., 2004), and that the phenomenon of epigenetic switching can arise as an adaptation to fluctuating selection during laboratory evolution (Beaumont et al., 2009). Numerous theoretical models have also been developed (Jablonka et al., 1995; Lachmann and Jablonka, 1996; Thattai and Oudenaarden, 2004; Kussell et al., 2005; Kussell and Leibler, 2005; Wolf et al., 2005; Salathé et al., 2009; Visco et al., 2010; Gaál et al., 2010; Libby and Rainey, 2011), all of them concluding that epigenetic switches can confer an adaptive advantage under specific circumstances (see Section 3.2 and Appendix C). Nevertheless, these models assumed a static (non-evolving) population with two discrete phenotypes, and evaluated the fitness advantage over a set of switching rates. These approaches ignored the question of how the epigenetic switches first evolved –its adaptive origin–, as well as the competition with universal genetic adaptation, and the effect of the population dynamics.

In order to explore the conditions that select for epigenetic switching and the extent to which epigenetic switches outcompete adaptation through genetic mutation in fluctuating environments, I have used computer simulation to evolve a mechanistic model of a self-activating genetic circuit which can adapt genetically and exhibit epigenetic switching. In this chapter, I will recapitulate the different strategies that organisms can employ to deal with fluctuating environments, which were briefly introduced in the Introduction (see Section 1.7), and emphasize the costs and advantages of each of these strategies (Section 3.1). I will then discuss previous work on the role of epigenetic switches and other mechanisms in fluctuating environments, in a similar conceptual framework to the one used in my project (Section 3.2). Afterwards, I will describe my evolutionary and mechanistic model and the series of computational experiments using this model (Section 3.3), as well as alternative assumptions that were tested in the model (Section 3.4). Finally, I will discuss the model limitations and future improvements (Section 3.5).
3.1 Strategies to cope with fluctuating environments

In general, organisms can apply very diverse mechanisms to deal with fluctuating environments. An important question in evolutionary biology is when and how each of these mechanisms are selected and used. For simplicity, I will focus on a particular scenario where only two discrete environmental states exist, and fluctuations between these occur on the timescale of generations. This is a complicated system with multiple factors and dimensions, and consequently the problem has been approached from several perspectives. Importantly, every theoretical model has overlooked some aspect of it. The response strategies analyzed in theoretical papers can be classified into four general groups: genetic adaptation (through mutations), stochastic epigenetic switching (bet-hedging through a bistable switch), inducible switching (sense and respond to the environment), and ignoring the environmental changes. In reality, a natural system can have properties from more than one of these groups. The simplified classification can help us to understand their properties and why each of them can be selected.

3.1.1 Genetic adaptation

Genetic adaptation refers to the accumulation of mutations in the population until a fitter phenotype is discovered, which will then be selected through natural selection. In theoretical or computational approaches, this strategy requires the consideration of a mechanistic model, where the genotype –with its respective phenotype– can be gradually changed, and both beneficial and non-beneficial mutations can occur. The only previous work that takes this relevant mechanism into account is Kuwahara and Soyer (2012) (see Section 3.2.2).
3.1.2 Stochastic epigenetic switching

The stochastic epigenetic switch consists of a bistable system disconnected from the environment, where two possible phenotypic states exist with a stochastic transition frequency due to biochemical noise. As described in the introduction (see Section 1.4.2), the transition frequency depends directly on the regulatory parameters that affect biochemical noise and the stability of each state, which can be tuned by natural selection. In a mechanistic model, these transitions can be driven by the biochemical noise (Ribeiro, 2008; Kuwahara and Soyer, 2012). This mechanism corresponds to a bet-hedging strategy when the transitions last for multiple generations (i.e. the phenotypic state can be inherited).

3.1.3 Inducible switching

Inducible switching includes a wide spectrum of sensing accuracy and efficiency on the phenotypic transition (Wolf et al., 2005). In the extreme case, where the system is highly accurate and efficient, the phenomenon is called phenotypic plasticity (Jablonka et al., 1995; Meyers and Bull, 2002) (see Section 1.7.1). More generally, phenotypic plasticity could be thought as a bistable system where the environment works as the stimulus that favors –in the stochastic view– or induces –in the deterministic view– the switching between states. Previous theoretical work has shown that an inducible system is preferred when: the cost of the sensing mechanism is low, the time delay after induction is short, and good informational cues exist. I have not explored phenotypic plasticity in my thesis. Even so, the previous work on phenotypic plasticity are discussed in the next section, and the potential of extending our model to include environmental sensing is considered in Chapter 6.
3.1.4 Ignoring the environment

Ignoring the environmental changes simply means that the population stays in the “original” state regardless of the changes in the environment. Several previous theoretical works assumed that the population can use this as a strategy to deal with fluctuating environments: if the environmental change is infrequent and transitory, the long-term growth rate can be maximized if no attempt to adapt is made (e.g.: Wolf et al. (2005), Salathé et al. (2009), and Gaál et al. (2010)).

3.2 Review of previous work on the selective advantage of epigenetic switches

3.2.1 Assuming a static, non-evolving population

Almost all previous theoretical works have assumed a static, non-evolving population, considering only the long-term growth rates of the population (Jablonka et al., 1995; Lachmann and Jablonka, 1996; Thattai and Oudenaarden, 2004; Kussell et al., 2005; Kussell and Leibler, 2005; Wolf et al., 2005; Gaál et al., 2010). In general, the authors assumed that a finite number of phenotypes existed, each optimal in a different environment, and then compared the different switching rates, which could depend on the environment.

For instance, Jablonka et al. (1995) compared three different strategies: non-inducible switching (a stochastic epigenetic switch with a small transition rate), completely inducible switching, and an intermediate response (an inducible bistable switch with memory). They considered both the induction delay and the phenotypic memory as tunable properties. They observed that the intermediate response is advantageous under random environmental fluctuations; if the environment is strictly periodic, the inducible system is favored unless fluctuations occur faster than the induction delay.
Lachmann and Jablonka (1996) explored the optimal values for the transition rates under fluctuating environments. The authors concluded that for stochastic epigenetic switches, the optimal rate for random transitions is around the frequency of the environmental fluctuations. Analogously, Thattai and Oudenaarden (2004) considered that the transitions between phenotypic states depended on the environmental state, and explored under which circumstances a transition rate to the “unfit” state different to zero will be selected. They concluded that if the transition to the “fit” state is fast enough –short induction delay–, a homogeneous (i.e. zero transition rate to the “unfit” state) population will be always favored.

Similarly, Kussell and Leibler (2005) compared inducible to stochastic transitions, but took into account in account the cost of sensing, the induction delay and the diversity cost imposed by the stochastic switching. They concluded that a sensor is advantageous only if the environment is highly uncertain, and the stochastic switching will be favored when environments change infrequently. On a second paper, Kussell et al. (2005) considered only stochastic transitions, and observed that the type of environmental changes determines the optimal transition rate.

Wolf et al. (2005) considered more flexible adaptation strategies: going from ignoring the environment, a deterministic inducible response, stochastic inducible response, and pure stochastic switching. If no sensor exists, the detection of the sensor is bad, or long induction delays exist, stochastic switching is always selected under the time-varying environmental conditions tested. Gaál et al. (2010) explored the extent of these conclusions under asymmetrical environments. They observed that as the asymmetry increases, the selected strategy goes from the optimal stochastic switching population (where the transition rate is assumed equal in both directions) to an equally optimal non-switching and switching populations, to finally being optimal to ignore the environment. This occurs even if a local maximum still exists for a switching rate distinct to zero. Salathé et al. (2009) used a slightly different
approach to explore the impact of asymmetric fitness landscapes, assuming an infinite population, and following subpopulation frequencies through generations. They concluded that with an asymmetry penalty over a certain threshold –unless the selection pressure is very strong in both environments–, ignoring the environment becomes optimal over stochastic switching (with an optimal rate approximately equal to the environmental fluctuation frequency). Finally, Liberman et al. (2011) took Salathé et al. (2009) and Gaál et al. (2010) one step forward by including recombination in the model, and observed that recombination made unlikely that a stable non-zero transition rate exists.

A recurrent model to explore adaptation to fluctuating environments is the modifier locus model, where a major locus is assumed to determine the mean phenotype, and the modifier locus controls the variability on the major locus (otherwise selectively neutral). In some previous work, this modifier locus has been presented as a mutator, determining the frequency of mutations or switching of alleles (i.e. alternative sequences of a locus) in the major locus. In other words, the authors presented this model as a genetic adaptation strategy (e.g. Salathé et al. (2009), and Liberman et al. (2011)). Nevertheless, in all these cases, a small number of alleles was considered (e.g. 2) and the rate of mutation was usually high in both directions. For this reason, I regard these examples as using stochastic switching strategy, rather than epigenetic adaptation (see Appendix C).

Carja et al. (2013) used the modifier model to explore the selection of the phenotypic variance over generations in a fluctuating environment. The authors showed that the interaction of stochastic epigenetic variation and recombination in a fluctuating environment can allow a genotype with a lower geometric mean fitness to fix in the population. In Furrow and Feldman (2014), the modifier locus role was explicitly defined as inducing an epigenetic modification in the major locus, and the model was expanded to consider inducible switching. The authors concluded that both
the underlying genetic variation and the dynamics of the fluctuating environment determined the selection of epigenetic mechanisms. Finally, Carja et al. (2014b) and Carja et al. (2014a) expanded the modifier model even more to explore the effect of other sources of variation: stochastic switching, recombination and migration. The authors observed that these three evolutionary forces responded similarly to fluctuating environments.

Distinct environmental scenarios can also be considered, leading to somewhat different observations. For instance, the environmental fluctuations can occur periodically or randomly, with symmetrical or asymmetrical propensities or rates, or as rare, occasional catastrophic events. Visco et al. (2010) explored the selection of stochastic epigenetic switching under a single environment with occasional and instantaneous catastrophic events, whose rate depended on the population fitness. They observed that the stochastic epigenetic switching strategy was favored by strong catastrophes, while a non-switching strategy (i.e. ignoring the environment) by weak catastrophes. Libby and Rainey (2011) considered a strong frequency-dependent selection (i.e. the fitness of each genotype depends on its frequency in the population), with an exclusion rule for the subpopulation with the highest fitness (i.e. the most frequent genotype was excluded in the next generation) and bottleneck (i.e. only a small fraction of the population passed to the next generation) when the environment changes. Even considering a switching cost –reducing the growth rate on switching genotypes–, exclusion rules favored switching phenotypes; on the other hand, larger (weaker) bottlenecks permitted faster-growing, non-switching types to pass through to the next “round” thereby outgrowing the switching type.

3.2.2 A mechanistic model

As mentioned above, considering a mechanistic model allows for different questions to be address. First, the role of the biochemical noise in the stochastic transition rate
can be explored. For example, Ribeiro (2008) modeled individual cells as bistable switches and explored the population behavior under a fluctuating environment, considering both inducible and stochastic epigenetic switching. The author concluded that the optimal noise level depends on the environmental fluctuation frequency, with higher noise being selected in fast fluctuating environments. In bistable systems, the stochastic transition rate is directly dependent on the biochemical noise, which allows noisier genotypes to adapt faster to the environmental fluctuations.

A mechanistic model where genetic adaptation can be implemented is necessary to study the adaptive origin of any of these strategies. Specifically, two aspects of the model must be considered: (1) which mechanisms and phenotypes are actually available, and (2) what is their fitness cost. A self-regulated gene circuit is the simplest mechanistic model that can exhibit either monostable gene expression (i.e. one stable state) or bistable gene expression (i.e. two stable states) (see Section 1.4.2). The phenotype (“gene expression”) under selection depends on the underlying kinetic parameter set (“genotype”), such as protein synthesis, protein degradation, protein-protein and protein-DNA interaction rates.

Kuwahara and Soyer (2012) were the first exploring de novo selection of stochastic epigenetic switching in a fluctuating environment. The authors developed a mechanistic model capable of displaying bistability, and analyzed the evolution of bistability and noise starting from monostable genotypes. The authors observed that populations can adapt to the fluctuating environment, and the level of fitness improvement depends on the particular evolutionary conditions (e.g. mutation rate and environmental fluctuation frequency). All their simulations evolved high non-linearity values, both assuming stochastic and deterministic gene expression, and regardless of the eventual emergence of bistability. In all cases, they observed a decrease on the adaptation time compared to the initial conditions. They showed that the presence of noise (i.e. stochastic simulations) was beneficial with respect
to deterministic simulations only when stochastic switching was the best strategy to cope with the environmental fluctuations. Otherwise, the stochasticity imposes a burden in the population and the adaptation process. Additionally, they observed a significant increase in phenotypic diversity in the population right after each environmental change, suggesting that an increase in noise can be advantageous under these circumstances. Finally, they detected the emergence and maintenance of bistability under a broad range of environmental fluctuation and mutation rates, but only in the presence of noise.

Nevertheless, Kuwahara and Soyer (2012) proposed that bistability emerged only as a byproduct of the selection for increased nonlinearity, which is beneficial for an increased mutational and phenotypic diversity, resulting in higher evolvability. There are two faults with their conclusions. First, a bistable system strictly requires a high nonlinearity; and if this bistable system is a possible solution for the environmental dynamics, it would have by definition a high “evolvability”, given the used metric (i.e. the ratio of the fitness improvement over the extent of genetic change, which would be zero or very small when epigenetic switching strategy is implemented). Consequently, a stochastic switch will not be selected without these properties, and the line of causality cannot be determined without further analysis. Second, the authors claimed that increasing the nonlinearity results in higher noise, and that this was beneficial; but this trend does not hold true when the system reached the bistable region (see Section 5.2.1).

3.2.3 Experimental approaches

There have been several interesting experimental approaches to exploring adaptation through fluctuating environments using epigenetic switches. For example, Balaban et al. (2004) used a microfluidic device to study and quantitatively characterize the bacterial persistence phenomenon, supporting the bet-hedging hypothesis previously
proposed. Beaumont et al. (2009) took this one step forward and proved that \textit{de novo} evolution of bet-hedging can be observed in experimental bacterial populations if the right selection pressures are applied (e.g. exclusion rule plus bottleneck). This is later more deeply analyzed (theoretically) by Libby and Rainey (2011) who showed the role of the bottleneck strength and the exclusion rule efficiency in the selection of stochastic switchers.

3.3 Our approach – A computational experiment

Briefly, I have used computational simulations to evolve a self-activating gene, the simplest genetic circuit that can apply either genetic adaptation or stochastic epigenetic switching strategies, in a population of cells surviving in a fluctuating environment. Each generation, I simulated the stochastic gene expression (i.e. phenotype) for each cell with a set of biophysical parameters (i.e. genotype). I modelled the evolutionary process for a population of size $N$ using a modified Wright-Fisher model with selection and mutation. Cells with phenotypes that best matched the current environment were preferentially selected to populate the next generation. Each selected cell could mutate its genotype with a fixed probability $u$. The environment fluctuated with frequency $\nu$ between two discrete states where each environment selected a different gene expression level. The evolutionary dynamics of the population were analyzed over a wide range of selection pressures, mutation rates, mutation step-sizes, population sizes, and environmental fluctuation frequencies. Each condition tested corresponds to an independent \textit{computational experiment} and the results can be easily analyzed and compared.

As described in the previous section, Kuwahara and Soyer (2012) implemented a similar model to simulate the evolution of a self-activating gene in a fluctuating environment. Important differences exist between our models. For instance, Kuwahara and Soyer (2012) simulated the mRNA and protein dynamics, while in our model
the transcription step is assumed in quasi-equilibrium and omitted. They assumed that cell division reduced the phenotype value, dividing the parental phenotype by half in the cloned cell in the next generation; however, we considered that even if cell division is expected to introduce noise in the system, both by reducing the number of molecules and by the randomness of the process, the volume will also be divided by half, and the reaction rates will scale consequently. Also, Kuwahara and Soyer (2012) used a selection scheme where selection strength is not tunable, so they could not explore the effect of this parameter. We argue that a unimodal fitness distribution (e.g. Lorentzian function) is more biologically realistic than their choice of a sigmoidal fitness. Analogously, we believe that unlike additive mutation –as adopted in their simulations–, multiplicative mutation better reflects how mutations affect thermodynamics of protein stability, protein-DNA and protein-protein interactions. Besides, the mutation step-size is fixed in their model and its effect was not considered. Finally, I have explored a significantly higher number of evolutionary conditions, including some cases where their conclusions clearly do not hold.

A self-activating gene can display a unimodal expression distribution (monostable) or a bimodal expression distribution of protein number (bistable), depending on the underlying biophysical parameters (Figure 3.1A-B). The spontaneous transitions (i.e. epimutation) between bistable phenotypes are driven by the stochasticity on gene expression, i.e. the biochemical noise. The genetic mutations change the biophysical parameters (i.e. genotype) and modify both the distribution of protein number (i.e. phenotype, $\rho(A)$) and the stochastic transition rates between bistable phenotypes.

Two qualitatively different strategies of evolutionary adaptation could emerge from a self-activating gene (Figure 3.1C). The population could evolve from one monostable distribution to another by mutating its biophysical parameters after each environmental change (i.e. genetic adaptation). Alternatively, the population could
reside at a bistable solution where each bimodal state is optimal in one of the environments and epimutations from one bistable state to another occur over time without an underlying genetic mutation (i.e. epigenetic switching). In both cases, the phenotypic distribution of the population $P(A)$ expands each generation due to gene expression noise and mutations, but natural selection keeps it centered on the optimal protein number as determined by the fitness function of the current environment. After an environmental change, the fitness function changes and the tail of the phenotypic distribution with higher fitness will be selected every generation. In the case of genetic adaptation, this selection will gradually shift the population towards the new optimal phenotype through the accumulation of new mutations until the population is well-adapted again. The speed of genetic adaptation depends on the rate of arrival of fitter mutations, as well as the selection pressure. When the population applies epigenetic switching, the phenotypic noise also includes epimutations, which increase the maladapted fraction in the population (epimutational load). However, after the environment and the fitness function change, the “epimutated” individuals rapidly overtake the population, quickly shifting the distribution to the new optimal value. Thus, genetic adaptation and epigenetic switching can directly compete as two strategies of adaptation to a fluctuating environment.

3.3.1 Mechanistic model

For simplicity, I considered only two biochemical events that either increase –synthesis– or decrease –degradation– the number of proteins $A$ in the cell by one molecule:

$$\text{Protein synthesis : } A \xrightarrow{f(A)} A + 1 \quad (3.1)$$

$$\text{Protein degradation : } A \xrightarrow{\gamma A} A - 1 \quad (3.2)$$
Figure 3.1: Genetic adaptation versus epigenetic switching of a self-activating gene. (A) Diagram of a self-activating gene that considers two biochemical events: protein synthesis with a rate that increases with number of proteins \( A \) (i.e., positive feedback loop) and protein degradation. On the right, a cartoon of the biophysical parameter space or genotypes (\( \theta \)) of this gene circuit with two characteristic regions: monostable (white) and bistable (pink) phenotypes. In the monostable region, a genotype might be optimal in either one environment (e.g. LOW protein numbers, \( \theta_L \)) or the other (e.g. HIGH protein numbers, \( \theta_H \)). Genetic mutations are required to change from one solution the other (blue arrow). In the bistable region, a single genotype (e.g. \( \theta_B \)) can display two different phenotypes with each phenotype potentially optimal in both environments. (B) Cartoon of the protein number (\( A \)) dynamics in an individual cell with each of the genotypes described in (A). Monostable genotypes (\( \theta_L \) and \( \theta_H \)) exhibit a unimodal distribution of protein expression (\( \rho(A) \)), whereas a bistable genotype (\( \theta_B \)) exhibits a bimodal distribution having spontaneous transitions between phenotypic states over time (i.e. epimutations) triggered by stochastic gene expression. Cartoon of the population dynamics using (C) genetic adaptation or (D) epigenetic switching to adapt after an environmental change. The fitness score function (\( \omega \); orange dashed line) and the phenotype distribution of the population (\( P(A) \); blue line) are shown for each generation (\( g \)), and the fraction of the population expected to be selected in the next generation (i.e. individuals with higher fitness scores) are highlighted (blue area). The environment changes from selecting HIGH protein numbers (light green) to select LOW protein numbers (dark green).
The synthesis rate $f(A)$ describes the probability per unit time that protein synthesis occurs, and it is a nonlinear function of activator $A$:

$$f(A) = k \cdot \left( \alpha + (1 - \alpha) \frac{A^{n_H}}{A^{n_H} + K_D^{n_H}} \right) \quad (3.3)$$

where $k$ (proteins/unit time) represent the maximum synthesis rate, $\alpha$ is the basal synthesis rate relative to $k$, $K_D$ (proteins) is related to the protein-DNA dissociation constant, and $n_H$ is the degree of molecular cooperativity (i.e. Hill coefficient). The degradation rate $\gamma \cdot A$, which describes the probability per unit time that protein degradation occurs and that $A$ is decreased by one, is a linear function of activator $A$ where $\gamma$ (1/unit time) is the protein degradation rate constant.

**Steady state solutions.** The gene expression dynamics in this model can also be represented as an ordinary differential equation of the form:

$$\frac{\partial A}{\partial t} = f(A) - \gamma \cdot A \quad (3.4)$$

where the discreteness of the protein numbers is ignored, and deterministic dynamics are assumed. Still, no analytical solution can be obtained. Nevertheless, the deterministic steady states can be determined numerically for any given set of biophysical parameter values (see Appendix D).

As mentioned above, this system displays bistability for some biophysical parameter values. When the biophysical parameter values lay in the bistable region, the $f(A)$ function intersects three times the degradation rate function $\gamma \cdot A$, creating two stable steady states ($A_L^*$ and $A_H^*$) with an unstable steady state in the middle ($A_U^*$). A useful analogy to understand these steady states is the potential energy $V(A)$, which is defined by $\partial V / \partial A = -\partial A / \partial t$ (see Section 1.4.2). In the potential energy landscape, the two local minima correspond to the stable steady state $A_L^*$ and $A_H^*$; then, these steady states work as system attractors, where after any small pertur-
Figure 3.2: Monostable and bistable solutions. Different parametrizations of the model are shown, one resulting in a monostable solution (left), and other one displaying bistability (right). The synthesis (green line) and degradation (gray dashed line) rate functions are shown; whenever these two functions intersect, a steady state occurs. The corresponding potential energy function $V(A) = -\int_0^a (f(a) - \gamma \cdot a) da$ is shown in the bottom. The stable steady states appear as basins in $V(A)$, and work as attractors (i.e. after any small perturbation, the system is attracted back to the steady state); while the unstable steady state in the bistable example corresponds to a hill in the $V(A)$ function, and any small perturbation will pull the system to one of the attractors. As a reference, the arrows show the direction of the change on $A$ (i.e. $\partial A/\partial t$ sign) around the steady states.

bation, the system will be attracted back to the current steady state. On the other hand, the local maximum corresponds to an unstable steady state $A^*_U$, where any infinitesimal perturbation will push the system to another steady state. When the system is monostable, the $f(A)$ function intersects only once the degradation rate function $\gamma \cdot A$, creating a unique minimum in the potential energy function $V(A)$, which will work as an attractor (i.e. a stable steady state; Figure 3.2).

Stationary distribution. Once the discreteness of molecule numbers is taken in account, the system dynamics is described by the Chemical Master Equation (CME), which describes the change of the probability $P(A, t)$ to have $A$ proteins (i.e. our system state) with time $t$ (see Section 2.3.1 for a detailed explanation of the CME).
In this system, the CME is:

$$\frac{dP(A,t)}{dt} = \sum_k (f(A - 1) \cdot P(A - 1, t) + \gamma \cdot (A + 1) \cdot P(A + 1, t) - (f(A) + \gamma \cdot A) \cdot P(A, t)),$$

(3.5)

where $A \in \{0, 1, 2, \ldots\}$. This equation can also be represented in a matrix form:

$$\frac{dP(A, t)}{dt} = B \cdot P(A, t),$$

(3.6)

where $A = [0, 1, 2, \ldots]^T$ is the vector of all possible states (i.e. the protein number space) and $B$ is the state transition matrix such that

$$B_{ij} = \begin{cases} -(f(A_i) + \gamma \cdot A_i) & \forall i = j \\ \gamma \cdot A_i & \forall A_j = A_i - 1 \\ f(A_i) & \forall A_j = A_i + 1 \\ 0 & \text{otherwise.} \end{cases}$$

(3.7)

This system is irreducible, given that each biochemical event changes the protein number $A$ by $\pm 1$, and then every possible state is reachable from any other possible state by a finite number of reactions. Zeron and Santillán (2011) has shown that any irreducible chemical reaction system has a unique and asymptotically stable stationary distribution $P^*$ such that

$$\frac{dP^*(A, t)}{dt} = B \cdot P^*(A, t) = 0,$$

(3.8)

and this corresponds to the nonzero eigenvector $\Lambda \geq 0$ in the kernel (i.e. nullspace, or set of eigenvectors with eigenvalue $\lambda$ zero) of $B$ after renormalization:

$$P^*(A_i) = \frac{\Lambda_i}{\sum_j \Lambda_j} \geq 0$$

(3.9)

where $\Lambda_i$ is the $i$th element in the vector $\Lambda = [\Lambda_0, \Lambda_1, \Lambda_2, \ldots]^T$, and then $\sum_i P(A_i) = 1$. Then, if an upper limit ($A_{MAX}$) is assumed, such that $P(A > A_{MAX}, t) = 87$
A monostable system can display bimodality. An example showing a bimodal protein number stationary distribution $P^*(A)$ (the local maxima are shown as black circles), even if only one steady state solution (i.e. $\dot{A}/\dot{t} = 0$) exists.

0 $\forall t$, the stationary distribution in our system can be estimated for any given biophysical parameter values by finding numerically the eigenvector $\Lambda$. I used the Arnoldi iteration numerical method to estimate $\Lambda$ given that our transformation matrix is not symmetric (Lehoucq and Sorensen, 1996).

**Bimodality versus bistability.** Bistability is only strictly defined in the deterministic world. When the stochasticity in the system and the resulting stationary distribution are considered, bimodality is often used as a simile of bistability. Nevertheless, it is possible to obtain bimodality by a deceleration in the flux towards a unique attractor in the system (for example, see Figure 3.3). In our analysis, a system is defined as bistable if and only if its deterministic solution has two stable steady states, not whether its stationary distribution is bimodal.

**Stochastic switching rate.** Given a bistable system, we can estimate the frequency of stochastic switching between phenotypic states or modes. For this I used the mean passage time to estimate the stochastic switching rate for an individual cell to hop from the LOW to HIGH expression level ($1/\tau_{L\rightarrow H}$) and *vice versa* ($1/\tau_{H\rightarrow L}$), defining these states as being under (LOW) or over (HIGH) the unstable steady state ($A^*_U$),

![Figure 3.3](image-url)
also called saddle point. I assumed that once the system crosses the saddle point between the stable states, the epigenetic transition has occurred and that the region around the current steady state is in quasi-equilibrium (Roma et al., 2005). If no reverse transition occurs in the same generation, the expected number of epigenetic switching events \( T \) per generation is \( N \cdot t_{\text{MAX}} / \tau \), where \( N \) is the population size and \( t_{\text{MAX}} \) the individual’s life span.

Our system is a continuous-time Markov chain with a finite (discrete) state space (under the assumption that \( \text{Prob}(A > A_{\text{MAX}}) = 0 \)), and time homogeneous (i.e. reaction rates are time independent). Then, the mean passage time to the saddle point \( (A_U^*) \) starting at state \( A_0 \) is defined as \( b(A_0) = \mathbb{E}[\inf\{t : A_t = A_U^*\}] \), where \( A_t \) is the state or protein number at time \( t \). Then, considering all possible initial states, we obtained

\[
\bar{b} = [-\bar{B}]^{-1} \cdot \bar{1}
\]

(3.10)

where \( \bar{B} \) is the state transition matrix excluding the row and column associated to the state \( A_U^* \) (Lawler, 2006). Then, the expected time from the system to go from the first to the second basin of attraction in the potential energy landscape \( (V(A)) \) is:

\[
\tau_{L \rightarrow H} = \sum_{x=0}^{[A_U^*]} b(x) \cdot P_x^*
\]

(3.11)

\[
P_x^* = \frac{P^*(A = x)}{\sum_{i=0}^{[A_U^*]} P^*(A_i)}
\]

(3.12)

assuming the system is in quasi-equilibrium in the first basin. Then, every generation \( N \cdot t_{\text{MAX}} / \tau_{L \rightarrow H} \) individuals are expected to have a stochastic transition from the LOW state to the HIGH state, where \( N \) is the population size and \( t_{\text{MAX}} \) is the individual’s life-span. Analogously, the expected time from the system to go from the second to
the first basin of attraction in the potential energy landscape \((V(A))\) is:

\[
\tau_{H \rightarrow L} = \sum_{y = [A^*_y]}^{A_{MAX}} b(y) \cdot P^*_y (3.13)
\]

\[
P^*_y = \frac{P^*(A = y)}{\sum_{i = [A^*_y]}^{A_{MAX}} P^*(A_i)} (3.14)
\]

assuming the system is in quasi-equilibrium in the second basin. Then, every generation \(N \cdot t_{MAX}/\tau_{H \rightarrow L}\) individuals are expected to have a stochastic transition from the HIGH state to the LOW state. For this we assume that \(t_{MAX}/\tau_{L \rightarrow H} \ll 1\) and \(t_{MAX}/\tau_{H \rightarrow L} \ll 1\), i.e. no more than one transition is expected to occur per individual in its life-span in either direction; if this assumption is not satisfied, a more complex probabilistic model needs to be considered.

**Biophysical parameter space & the phenotypic landscape.** To gain some intuition about the effect of each parameter over the system’s solution, I analyzed the limit where the nonlinearity goes to infinity \((n_H \rightarrow \infty)\). In this extreme case, the synthesis rate becomes a discontinuous function of the form:

\[
f_{n_H \rightarrow \infty}(A) = \begin{cases} 
\alpha k & \text{if } A < K_D \\
\frac{1+\alpha}{2} k & \text{if } A = K_D \\
k & \text{if } A > K_D 
\end{cases} (3.15)
\]

where \(\alpha \in [0, 1]\), \(k, K_D \in \mathbb{R}_+\), and \(A \in \mathbb{R}_{\geq 0}\). Then, if \(K_D < \alpha k/\gamma\), the system is monostable with the steady state \(A^*_L = k/\gamma\); if \(k/\gamma < K_D\), the system is monostable with steady state \(A^* = \alpha k/\gamma\) (Figure 3.4A). This system is bistable if \(\alpha k/\gamma < K_D < k/\gamma\), with steady states \(A^*_L = \alpha k/\gamma\) and \(A^*_H = k/\gamma\); which steady state is reached depends of the initial protein value \((A_0)\). If \(A_0 < K_D\), the system is attracted to the \(A^*_L\) steady state, if \(A_0 > K_D\), to the \(A^*_H\) steady state, and if \(A_0 = K_D\), it depends on the sign of \(\frac{1+\alpha}{2} k - \gamma \cdot K_D\), going to \(A^*_L\) if negative and to \(A^*_H\) if positive. Also, in the specific case that \(\frac{1+\alpha}{2} k - \gamma \cdot K_D = 0\), an unstable steady state appears at \(A^*_U = K_D\).
Figure 3.4: Biophysical parameters, deterministic steady state solutions and stochastic stationary distributions of protein levels. (A) The effect of the maximum synthesis rate ($k$) and the affinity constant ($K_D$) over the deterministic steady state solutions of the protein expression (i.e. $\frac{dA}{dt} = f(A^*) - A^* = 0 \Leftrightarrow f(A) = k(\alpha + (1 - \alpha)\frac{A^{n_H}k}{A^{n_H} + K_D^{n_H}})$) in the limit of high Hill coefficients ($n_H \to \infty$). If $K_D < \alpha k$ the system is monostable HIGH with the protein expression steady state ($A^*$) equal to $k$; on the other hand, if $K_D > k$ then the system is monostable LOW with $A^* = \alpha k$. When $\alpha k \leq K_D \leq k$ is intermediate, these two steady states coexist and the system is bistable. (B) Bifurcation diagram of the protein steady states as the Hill coefficient ($n_H$) varies while keeping the rest of the biophysical parameters fixed. As $n_H$ value increases, the system goes from monostable (blue dots) to bistable (violet and pink dots). As $n_H \to \infty$, the stable steady states monotonically approach their limiting values, $\alpha k$ and $k$ (dashed gray lines), and the unstable steady state asymptotically approaches $K_D$ (dotted gray line). I show a few examples of the stationary distribution of the protein expression for stochastic simulations with intrinsic biochemical noise (bottom). As $n_H$ approaches the bifurcation point (where the system passes from being monostable to bistable) the stationary distribution becomes wider (i.e. the phenotype is more variable). In the bistable region, even if the two modes of the stationary distribution do not change much, their relative weights can be significantly affected by the value of the unstable steady state, as stochastic transitions from one stable mode to the other become more or less probable.
Varying the Hill coefficient $n_H$ can take the system from the bistable region to the monostable region by collapsing $A_L^*$ and $A_U^*$ (when $f(A)$ is tangent to $\gamma \cdot A$ in that point) and finally making these fixed points disappear (Figure 3.4B). As the $n_H$ value increases, the stable steady states approach their limit values ($A_L^* \rightarrow \alpha k/\gamma$; $A_U^* \rightarrow k/\gamma$) and the unstable steady $A_U^*$ state approaches $K_D$.

The maximum synthesis rate $k$ and the degradation rate $\gamma$ depend on time. With no loss of generality, I reduced the number of free parameters in our model by substituting time $t$ with a time-dimensionless variable $\tau = t \cdot \gamma$. Many proteins in bacteria are not actively degraded and are diluted through cell growth and division. Thus, $\tau$ and $k$ are in units of cell cycle time. All mutated parameters were constrained to lie within a physiological range ($10^{-2} \leq k \leq 10^3$, $10^{-2} \leq n_H \leq 16$, $10^{-2} \leq K_D \leq 120$) for a bacterium such as *Escherichia coli*. The number of molecules for a transcription factor ranges between $0 - 10^3$ proteins per bacterium or concentration range $0 - 10^3$ nM for a bacterial volume of 1 fL (Li et al., 2014; Milo and Phillips, 2015). The DNA dissociation constant ($K_D$) has a similar range to the underlying transcription factors (Buchler et al., 2003; Milo and Phillips, 2015).

### 3.3.2 Evolutionary model

For simplicity, I evolved a haploid, asexual population with non-overlapping generations in a fluctuating environment (Figure 3.5). The underlying biophysical parameters depend on protein stability, protein-protein, protein-RNA, and protein-DNA interactions, which can increase or decrease through genetic mutations. For simplicity, I allowed mutations on the maximum synthesis rate ($k$), Hill coefficient ($n_H$) and DNA dissociation constant ($K_D$) during our evolutionary simulations, while keeping basal activity ($\alpha = 0.25$) and degradation rate (rescaled, $\gamma = 1$) fixed. The set of variable parameters $\{k, n_H, K_D\}$ are the genotype ($\theta$). The environment fluctuated periodically with frequency $\nu$ between LOW (selects for optimal phenotype $A^{(L)} = 20$
Figure 3.5: Evolutionary model. The environment fluctuates periodically with frequency $\nu$. The total number of generations spent in a constant environment (epoch) has the same length ($1/\nu$) and each environment (HIGH or LOW) selects for a different distribution of protein levels (phenotypes). Each generation, I simulated the stochastic protein dynamics of a self-activating gene in each cell across a population of size $N$. At the end of each simulation, the population phenotypes varied because gene expression is stochastic and because cells can have different underlying biochemical parameters. The current environment in each generation assigned a fitness ($\omega$) to each cell based on its final protein level. I used tournament selection (where $s_t$ determines the strength of selection) to determine the next generation of cells according to their fitness. Each cell in the next generation was mutated with probability $u$, where the current set of biophysical parameters were multiplied or divided up to a maximum step-size of $M$.
its biophysical parameters (i.e. genotype) and initial protein level inherited from its parent in the previous generation for 4 units of time (the cell “life span”).

2. Evaluate the fitness of each cell $i$ based on the protein level at the end of its life span ($A_i$, phenotype). The individual fitness function ($\omega_i$) is:

$$\omega^{(E)}_i(A_i) = \frac{v^2}{v^2 + (A_i - A^{(E)})^2}$$

(3.16)

where $E = \{L, H\}$ is the current environment, $A^{(E)}$ is the optimal phenotype for each environment, and $v^2 = 0.2 \cdot A^{(E)}$ is the width of the Lorentzian function. I define the population fitness ($w$) as the average $\Sigma_{i=1}^N \omega^{(E)}_i(A_i)/N$ over all cells.

3. Select the next generation using Tournament selection (Goldberg and Deb, 1991; Blickle, 1996), where $s_t$ cells are chosen randomly from the population. The cell with highest fitness within the chosen cohort is cloned into the new population. This “tournament” is repeated $N$ times with replacement to create a new population. The tournament size ($s_t$) modulates the selection pressure, where small $s_t$ is weak selection (e.g. for $s_t = 1$, there is no selection pressure and only genetic drift because any randomly selected individual is the tournament winner). Increasing $s_t$ leads to stronger selection and a faster selective sweep of fitter cells (e.g. for $s_t = N$, only the fittest individual in the entire population will be cloned into the next generation).

4. Allow random mutations with a fixed probability ($u$) in each cloned cell. If a mutation occurs, the value of each parameter in the cell genotype is updated as follows:

$$k' \leftarrow k \cdot M^r \sin(\phi_1) \cos(\phi_2)$$

(3.17)

$$n_H' \leftarrow n_H \cdot M^r \sin(\phi_1) \sin(\phi_2)$$

(3.18)

$$K_D' \leftarrow K_D \cdot M^r \cos(\phi_1)$$

(3.19)

where $M$ is the maximum fold change (mutation step-size), $r \sim U(0, 1)$ is a uniformly
distributed random value between 0 and 1, $\phi_{1,2} \sim U(0, 2\pi)$ are uniformly distributed random values in radian units. Unlike additive mutation, multiplicative mutation better reflects how mutations affect thermodynamics of protein stability, protein-DNA and protein-protein interactions (Wells, 1990; Yoshikuni and Keasling, 2007). All mutated parameters were constrained to lie within a physiological range that is typical for a bacterial transcription factor (see Section 3.3.1).

5. If the evolutionary simulation is at the end of an epoch, then change to other environment; otherwise keep the same environment. Return to Step 1 to simulate the next generation.

**Gene expression dynamics.** The stochastic expression dynamics of our gene circuit were simulated using the Gillespie algorithm, which is a kinetic Monte Carlo method that explicitly simulates the probabilistic dynamics of a defined set of biochemical events (Gillespie, 1977). The propensity or probability rate ($r_j$) of chemical reaction $j$ occurring during the next interval $dt$ is related to the rates of mass-action chemical kinetics in a constant chemical reactor volume. In our simple biochemical network, the propensity of protein synthesis is $f(A)$ and the propensity of protein degradation is $\gamma \cdot A$. Each step, given the current number of chemical species, Gillespie’s direct algorithm first calculates the propensities $r_j$ and then calculates when the next reaction occurs and which one occurred. The waiting time of the next reaction is drawn from an exponential distribution with parameter $\Sigma_j r_j$, where the cumulative distribution function of any reaction occurring before time $\tau$ is $F(\tau) = 1 - e^{-(\Sigma_j r_j) \cdot \tau}$. Given that a reaction has occurred at time $\tau$, the probability that the event is reaction $i$ is equal to $r_i/(\Sigma_j r_j)$. Thus, for each iteration, two random numbers are required to determine $\tau$ and $i$ as drawn from the probability distributions. The random numbers are generated using the “Minimal” random number generator of Park and Miller with Bays-Durham shuffle and added safeguards (Press et al., 2002).
I also ran CONTROL simulations, where the biochemical noise in gene expression is excluded. I assumed the dynamic of the mean concentration (A) obeys a first-order ordinary differential equation (Eq. 3.4). The cell life span was assumed to be long enough for the steady state (i.e. \( \frac{\partial A}{\partial t} = 0 \)) to be reached before selection. I used numerical methods to calculate the steady state solution (phenotype) for any genotype and initial protein level inherited from the parent (Appendix D).

All Gillespie, deterministic, and evolution simulations were implemented in C++ (Appendix D), and all the analyses and figures were done using MATLAB.

### 3.3.3 Systematic test of evolutionary conditions

The in silico approach presented here enables a fast and efficient exploration of the evolutionary process in a fluctuating environment, permitting a systematic and extensive survey of different conditions. Given the well-controlled, reproducible conditions in these computational experiments, we have been able to uncover notable statistical regularities of the adaptation mechanisms to fluctuating environments at time scales that would impossible by traditional experimental approaches.

The population evolution was systematically tested under different evolutionary conditions varying the population size (N), the environmental fluctuation frequency (\( \nu \)), the selection pressure (s_t), the mutation rate (u), and mutation step-size (M) over a wide range of values (Table 3.1).

### 3.4 Alternative assumptions

I tested the robustness of our results to alternative choices and assumptions in the presented model by: changing the used evolutionary model (Section 3.4.1), allowing the environment to fluctuate randomly between the two possible states with mean frequency \( \nu \) (Section 3.4.2), using the average protein number or the distribution of protein numbers over the individual life span as phenotype (Section 3.4.3), changing
Table 3.1: Evolutionary parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Range</th>
<th>Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>$N$ Population size</td>
<td>${100, 250, 630, 1600, 4000, 10000}$</td>
<td>individuals</td>
</tr>
<tr>
<td>$\nu$ Environmental fluctuation frequency</td>
<td>${0.01, 0.02, 0.04, 0.0625, 0.10}$</td>
<td>$1$/generation</td>
</tr>
<tr>
<td>$s_t$ Selection pressure (i.e. tournament size)</td>
<td>${3 (47%), 6 (24%), 15 (10%), 40 (4%), 100 (1.7%), 250 (0.7%)}^a$</td>
<td>individuals</td>
</tr>
<tr>
<td>$u$ Mutation rate</td>
<td>${0.01, 0.03, 0.10}$</td>
<td>$(1$/individual) $(1$/generation)</td>
</tr>
<tr>
<td>$M$ Mutation step-size (i.e. maximum fold change)</td>
<td>${1.1, 1.4, 1.7, 2.1, 2.6, 3.2, 4.0, 5.0}$</td>
<td></td>
</tr>
</tbody>
</table>

^a The equivalence between tournament size and truncation selection is shown in parenthesis. The numbers in italics were inter- and extrapolated from the values presented in Blickle and Thiele (1995).

the fitness function to a Gaussian or a step-like function with similar span around the optimal phenotypes (Section 3.4.4), implementing different selection schemes (Section 3.4.5), as well as different mutation schemes (Section 3.4.6). Additionally, more quantitative aspects of the model were perturbed by exploring other optimal phenotypes for the environments (Section 3.4.7), basal activity $\alpha$ values (Section 3.4.8), and different degradation rate $\gamma$ values (Section 3.4.9). In addition, we allowed the basal activity $\alpha$ parameter to evolve simultaneously with $\{k, n_H, K_D\}$. The results of these alternative assumptions are discussed in the Chapter 5 and summarized in Figure 5.10.

3.4.1 Moran model

As described in the Introduction (Section 1.8), Wright-Fisher and Moran models are the most common options to simulate evolution. In our main simulations, I implemented a version of the Wright-Fisher model with non-overlapping generations. Alternatively, I tested a Moran model, where the reproduction and death events are treated as stochastic events allowing overlapping generations. At each time step, an individual is chosen for reproduction using the defined tournament selection scheme,
and an individual is randomly chosen from the population for death to keep the population size $N$ fixed. $N$ time steps occur in the previously defined lifespan time, such that the reproduction rate (and then mutation) is equivalent to the original model. Importantly, these simulations are considerably more expensive computationally, so I used a shorter simulation time (only 1,000 generations).

### 3.4.2 Environmental random fluctuations

The environmental fluctuations in our main simulations were regular and periodic with frequency $\nu$. I tested whether stochastic fluctuations with frequency $\nu$ produced different results; even though previous work demonstrated little difference between the two types of fluctuations (Jablonka et al., 1995; Thattai and Oudenaarden, 2004; Kussell and Leibler, 2005; Wolf et al., 2005). In these alternative simulations, the environment fluctuates randomly between the two possible states with mean frequency $\nu$.

### 3.4.3 Phenotype definition

Our simulations evaluated the protein number (phenotype) at the end of Gillespie simulation (individual life span) to calculate a fitness score given by a Lorentzian function centered the optimal phenotype. I also tested alternative phenotype definitions: (1) the life-time average protein number to assign its fitness score to each individual in the population, or (2) the life-time protein number distribution to calculate the average fitness score for each individual in the population.

### 3.4.4 Fitness functions

I also changed the shape of the fitness function from a Lorentzian to a Gaussian fitness function:

$$\omega_{g(E)}(A) = e^{-\frac{(A-A(E))^2}{2\sigma^2(E)}}$$  \hspace{1cm} (3.20)
where $\sigma^2_{(E)}$ is equal to the width in the Lorentzian fitness function ($v^2$); or a step-like function:

$$\omega^{(E)}_s(A) = \begin{cases} 1 & \text{if} \ (A - A^{(E)})^2 \leq 2\sigma^2_{(E)} \\ 0, & \text{otherwise} \end{cases} \tag{3.21}$$

where $\sigma^2_{(E)}$ is equal to the width in the Lorentzian fitness function ($v^2$).

### 3.4.5 Selection schemes

I used Tournament selection to select the next generation of cells based on the fitness of the individuals in the current generation. Other common selection schemes are Truncation, Proportional, and Weighted selection (Blickle and Thiele, 1995).

In the truncation selection scheme, only a certain fraction of the best individuals can be selected, each with the same probability. Blickle & Thiele (1995) calculated the truncation fraction that resulted in the same selection strength as a given tournament size (Table 3.1). I estimated that $s_t = 6$ corresponds to a 0.24 truncation fraction. I used this fraction in our Truncation selection simulation.

Using the proportional selection scheme, the probability of an individual to be selected is proportional to its fitness value. Similarly, in the weighted selection scheme, a random individual is picked from the population and is cloned into the new population if a uniformly distributed random number (from the interval [0,1]) is below its fitness. Importantly, the selection strength cannot be tuned in either of these two schemes.

### 3.4.6 Mutation scheme

Our simulations used a spherically symmetric 3D mutation scheme to permit covariation in biophysical parameters in a single mutational step. The mutation step size was determined by the radius of the spherical mutation, which was a uniformly distributed random value between 0 and 1 ($r \sim U(0,1)$). Such a radial density pro-
duces a non-uniform density of mutations with highest densities close to the parental phenotype because volume scales as $r^3$. I tested homogeneous spherical mutation by substituting $r$ in Eqs (3.17-3.19) with $\sqrt[3]{r}$ and a homogeneous cubic mutation where three uniformly distributed random value between -1 and 1 ($r_i \sim U(-1, 1)$) for each biophysical parameter. I also verified that mutating only one parameter at a time (1D mutation) and increasing the range of biophysical parameters to allow higher nonlinearity ($10^{-2} \leq n_H \leq 24$) and weaker DNA dissociation constants ($10^{-2} \leq K_D \leq 10^3$) did not fundamentally change our results.

3.4.7 Optimal phenotypes

The main simulations were performed with the LOW environment selecting for an optimal phenotype $A^{(L)} = 20$ proteins and HIGH environment for an optimal phenotype $A^{(H)} = 80$ proteins. The effects of doubling ($A^{(L)} = 40$ proteins, $A^{(H)} = 160$ proteins) and dividing by two ($A^{(L)} = 10$ proteins, $A^{(H)} = 40$ proteins) these values were explored.

3.4.8 Basal activity

At high levels of nonlinearity, the lowest protein level is $k \cdot \alpha$ and the highest protein level is $k$. A bistable, epigenetic switch has two solutions, each well-adapted to one of the environments only when the ratio $R = A^{(L)}_{opt}/A^{(H)}_{opt} = \alpha$ (Figure 3.4). Any mismatch between $\alpha$ and $R$ will disfavor epigenetic switching because an epimutation from an adapted mode will jump to a maladapted mode, after which the descendants must accumulate genetic mutations to further adapt. I explored the effect of other values of basal activity parameter ($\alpha = 0.2$, and $\alpha = 0.3$), but adjusting the LOW optimal phenotype accordingly ($A^{(L)} = 16$ proteins, and $A^{(L)} = 24$ proteins, respectively).

The rate of epimutation is sensitive to the frequency and magnitude of stochastic
events. The magnitude of stochastic events is inversely proportional to the total number of molecules. A higher rate of epimutation for smaller numbers of molecules is expected. The rate of epimutation should also increase as the two modes become closer. Thus, we expect a higher rate of epimutation for larger \( \alpha \).

3.4.9 Degradation rate

The protein degradation rate \( (\gamma) \) sets the timescale between stochastic events (i.e. faster protein degradation leads to more stochastic events per unit time during a Gillespie simulation). Thus, we expect a higher rate of epimutation for larger \( \gamma \).

3.5 Discussion & Conclusions

I verified that our observations were robust to many alternative model assumptions (Figure 5.10). Nevertheless, our simple model of a haploid, asexual population (Schifflers et al., 2011) omits some features of the evolutionary process. For example, variable population size, diploid genetics, sexual reproduction and linkage disequilibrium could all affect the evolutionary dynamics and selection of epigenetic switches in a fluctuating environment. Our model also fixed the mutation rate \( (u) \) and step-size \( (M) \) in each simulation, which imposes a mutational load when the population is adapted to a constant environment. Future simulations could allow natural selection to mutate and tune these parameters, which might favor genetic adaptation over epigenetic switching (Desai and Fisher, 2011). Our model also did not consider the case where mutations (e.g. adaptive mutation) or biophysical parameters (e.g. phenotypic plasticity) directly respond to changes in the environment. Last, I assumed that mutations continuously increase or decrease the biochemical parameters. This overlooked an important class of mutations, such as indels (i.e. rapid loss or abrupt change of function), gene duplication, and gene recruitment, which could abruptly change the topology of the gene network.
We considered the simplest genetic circuit that can exhibit epigenetic switching. However, alternative gene regulatory networks could generate different dynamics and phenotypes that are even better adapted to the fluctuating environment. For example, adding a negative feedback loop could reduce gene expression noise (Becskei and Serrano, 2000; Rosenfeld et al., 2002) or generate oscillations (Novak and Tyson, 2008; Tsai et al., 2008). An oscillatory gene circuit (e.g. circadian clock) might anticipate and respond to an environment that fluctuates regularly (e.g. day/night). Future research will explore more complicated gene regulatory circuits to understand the specific environmental dynamics and evolutionary conditions that favor oscillation versus epigenetic switching in the context of genetic adaptation. This should be of broad relevance to evolutionary biologists and systems biologists.
Populations adapt to environmental fluctuations

In order to evaluate the selective advantage of epigenetic switching over genetic adaptation, three aspects of the evolutionary process are essential: (1) solutions for the possible environments (i.e. phenotypes with a high fitness in each environment) must exist in the defined system; (2) these solutions must be accessible (i.e. mutations or epimutations should allow transitions between optimal phenotypes); and (3) both strategies of interest must have the potential to emerge (i.e. the system can display bistability). In this chapter, I show that our model satisfies these conditions. First, I characterize the population adaptation process to the environmental dynamics starting from a non-optimal genotype. I describe the forces driving the early adaptation, and I show this process is initially dominated by noise, and eventually the population moves to a high-nonlinearity regime (Section 4.1). This exemplifies that the “no-response” strategy proposed in previous works (see Section 3.1.4) is not a stable solution in our system (Section 4.1.1). Later, I characterize the population dynamics of optimal phenotypes/genotypes after each environmental change, emphasizing the distinct behaviors associated with the monostable and bistable populations (Section 4.2). Coexistence of monostable and bistable subpopulations was observed in almost
all conditions tested (Section 4.2.1). This observation was intriguing because the subpopulations applied distinct adaptation strategies: bistable subpopulations used epigenetic switching, while monostable subpopulations used genetic adaptation. I propose the constant “seeding” of one subpopulation to the other as the source of this coexistence (Section 4.2.2). Finally, I discuss these observations and the limitations of analyzing only temporal snapshots of the population (Section 4.3).

4.1 Adaptation to fluctuating environments

All populations initially started with a monostable genotype $\theta_0 = \{k = 80, n_H = 1, K_D = 10\}$ that was adapted to one environment but not the other (i.e. the steady state solution was $A^*(\theta_0) \approx 73$). These populations all evolved to a higher fitness solution and a strategy that was best suited for the underlying evolutionary parameters. For example, in a fast-fluctuating environment ($\nu = 0.1$) with small mutation step size ($M = 1.1$), the population initially had a good fitness only in the HIGH environment; after a few environmental cycles (an environmental cycle corresponds to a LOW epoch and a HIGH epoch) the population lost the advantage in the HIGH environment and display low fitness in both environments. Eventually, a dramatic increment on the geometric mean fitness per cycle $W_{cycle} = \frac{2^{\nu} \prod_{g=(cycle)} w_g}{\sqrt{\prod_{g=(cycle)} w_g}}$ occurred, and then the population was able to efficiently adapt in a few generations after each environmental change (Figure 4.1). I plot geometric fitness per cycle $W_{cycle}$ rather than fitness per generation ($w$) because it better reflects the long-term growth of fitter phenotypes (Orr, 2009).

The observed increment on fitness is related to the population migrating from a region of the genotypic space with low nonlinearity ($n_H \approx 1$), where the population tried to adapt after each environmental change by selecting mutations on the synthesis rate $k$, to a region with high nonlinearity ($n_H \approx 6$) inside the bistable region, where adaptation occurs through epigenetic switching (i.e. no genetic mutations
Figure 4.1: **Fitness as a population adapts to a fluctuating environment.** The initial population started from a non-optimal genotype ($\theta_0$) where $k = 80$, $n_H = 1$, $K_D = 10$ with evolutionary parameters $N = 10000$, $\nu = 0.1$, $s_t = 6$, $u = 0.03$, and $M = 1.1$. The geometric mean of the population fitness per environmental cycle ($W_{\text{cycle}}$) is shown, where each cycle spans a LOW (dark green) and HIGH (light green) epoch. The initial genotype was well adapted only to the HIGH environment, the population eventually evolved to a solution with higher $W_{\text{cycle}}$. This final genotype had high population fitness ($w$) in both environments and rapidly adapted after each environmental transition.

observed). Congruently, the population maximum synthesis rate $\langle k \rangle_g$ varied significantly during the cycle only when the population was in the monostable region with small $\langle n_H \rangle_g$ and $\langle K_D \rangle_g$ values. As $\langle n_H \rangle_g$ increases, the population quickly enters the bistable region and finds an optimal genotype, which shows little variation as the environment fluctuates (Figure 4.2). This transition occurred after approximately 1800 generations (i.e. $\sim 90$ cycles) in the shown example. Once the transition occurred, the population stayed in this global optimum in $W_{\text{cycle}}$ for the rest of the simulation (i.e. 10000 generations).

This example shows that in addition to adapting after each environmental change, the population adapted to the fluctuating environment by evolving to a region in the biophysical parameter space where the transition between solutions for each environmental state was efficient (Figure 4.2).
4.1.1 The “no-response” strategy is not a stable solution

In the example shown in Figures 4.1 and 4.2, the population had initially a slightly higher geometric mean fitness per cycle $W_{cycle} = \frac{2}{\sqrt{n}} \prod_{g=(cycle)} w_g$ because it accrued benefits by being adapted to the HIGH environment despite being maladapted to the LOW environment (Figure 4.3). During the LOW epoch that follows the HIGH epoch, the population shifted towards higher fitness values in the current environment ($w^{(L)}$). Nevertheless, the epoch was too short and mutation too weak for the population to perfectly adapt to the new environment before it changed again. The forces of mutation-selection in the alternative environment gradually shifted the population to an intermediate genotype that lay between the two optimal phenotypes for each environment. The population then evolved to the bistable parameter space with higher $\langle n_H \rangle_{cycle}$ and $W_{cycle}$. This final genotype had high population fitness ($w$) in both environments and rapidly adapted after each environmental transition.
through epigenetic switching (see inset).

Figure 4.3: Population fitness and nonlinearity per cycle. The initial population started from a non-optimal genotype ($\theta_0$) where $k = 80$, $n_H = 1$, $K_D = 10$ with evolutionary parameters $N = 10000$, $\nu = 0.1$, $s_t = 6$, $u = 0.03$, and $M = 1.1$. The geometric mean of the population fitness per environmental cycle ($W_{cycle}$) versus the average Hill coefficient per cycle ($\langle n_H \rangle_{cycle}$). Each cycle spans a LOW and HIGH epoch and there are 500 environmental cycles (increasing from gray to black) over 10,000 generations for this simulation. The insets show the population fitness per generation for the highlighted cycles (first, 89th, and last cycles), with the color bar in the top specifying the environmental state (LOW, dark green; HIGH, light green). The plot in the bottom is a close-up of the early cycles.

To better understand the forces that destabilize the “no-response” strategy, I performed a simple experiment by running the same simulation, but allowing only mutations in the maximum synthesis rate $k$ (i.e. $n_H = 1$ and $K_D = 10$ were fixed in the population). For simplicity and speed, a smaller population size $N = 100$
was used (Figure 4.4). Although selection tends to increase the average fitness each generation, this simple simulation shows that \( \langle k \rangle_g \) is surprisingly unconstrained and that \( W_{cycle} \) varies significantly over time (Figure 4.4A). To confirm that the observed behavior was not due to stochastic population dynamics or stochastic gene expression, an even simpler experiment was performed. I simulated a single Metropolis walker with deterministic gene expression dynamics (i.e. no biochemical noise; the steady state solution corresponds to the individual’s phenotype) in the same fluctuating fitness landscape (i.e. \( \nu = 0.1 \)), where a mutation (\( \theta' \)) occurs every generation (\( u = 1 \)) and it is accepted (i.e. \( \theta \leftarrow \theta' \)) with probability \( \min(1, w^{(E)}(\theta')/w^{(E)}(\theta)) \). I observed the same qualitative behavior than the previous simulation. The Metropolis walker simulation shows that the destabilization of the “no-response” strategy and unconstrained \( \langle k \rangle_g \) arises because of fluctuating fitness landscapes and not from genetic drift or gene expression noise.

I also repeated the full experiment with \( N = 100 \) and only \( k \) mutating, but starting with a higher fixed \( K_D \) value (\( K_D = 15 \) and \( K_D = 20 \)). The optimal \( k \) values for increasing \( K_D \) get shifted towards higher values, but the same qualitative behavior of the population is observed, where \( W_{cycle} \) varies significantly over time (Figure 4.5). Nevertheless, the average \( W_{cycle} \) gradually increased as \( K_D \) increased (Figure 4.5B). This observation explains the gradual shift of the population genotype towards higher \( K_D \) values in the early phase of evolution (Figure 4.2).

In conclusion, the evolutionary dynamics in early epochs were dominated by noisy genetic adaptation of a population maladapted to at least one of the environments, even if this implied decreasing \( W_{cycle} \) (Figure 4.4). The “no-response” behavior, i.e. being adapted to one environment and “ignoring” the alternative state, is not a stable solution for this system. Similar to previous work, this illustrates the importance of considering the full population dynamics in the adaptation process and not only the long-term average fitness (Cvijović et al., 2015).
Figure 4.4: Counteracting selection pressure destabilizes the “no-response” genotype in a fluctuating environment. (A) A population with an initial synthesis rate $k = 80$, and fixed Hill coefficient $n_H = 1$, and constant $K_D = 10$, was evolved with parameters $N = 100$, $\nu = 0.1$, $s_t = 6$, $u = 0.03$, and $M = 1.1$. The population average per generation $\langle k \rangle_g$ and population fitness $w$ (dark green for LOW, light green for HIGH) are shown; geometric mean fitness per cycle ($W_{cycle}$) is also shown (orange). A sample evolutionary trajectory over one cycle is shown in red, starting from the white-filled circle. (B) A single Metropolis walker with no biochemical noise in the same fluctuating fitness landscape, where a mutation ($\theta'$) occurs every generation ($u = 1$) and it is accepted (i.e. $\theta \leftarrow \theta'$) with probability $\min(1, w^{(E)}(\theta')/w^{(E)}(\theta))$. The maximum synthesis rate per generation $\langle k \rangle_g$ and individual’s fitness $w$ (dark green for LOW, light green for HIGH) are shown; geometric mean fitness per cycle ($W_{cycle}$) is also shown (orange). A sample evolutionary trajectory over one cycle is shown in red, starting from the white-filled circle.
**Figure 4.5:** Fitness advantage of increasing $K_D$ in low nonlinearity regimes. (A) Same approach and plot as in Figure 4.4A, but with a $K_D = 20$. (B) Comparing $W_{cycle}$ distribution for $N = 100$ populations with $n_H = 1$ and $K_D = \{10, 15, 20\}$, we observed that $W_{cycle}$ gradually increases as $K_D$ increases; the same simulation but in the bistable region ($n_H = 6; K_D = 46$) shows a significantly higher $W_{cycle}$.

### 4.2 Population dynamics during adaptation

To have a better insight of the adaptation process, I analyzed the population distribution for each generation after each environmental transition. The relation between the genotype and phenotype can be visualized by plotting the actual phenotype $A$ of each cell from stochastic simulation versus expected deterministic steady state value $A^*(\theta)$, which is determined by the individual’s genotype $\theta$. The two possible deterministic steady state values $A^*(\theta)$ are simultaneously shown for bistable individuals.
The population dynamics for the same example displayed in Figures 4.1-4.3 in the first environmental cycle, an intermediate cycle (89th), and the last cycle are shown in Figure 4.6. In the first and intermediate cycles, the population is completely monostable (i.e. the bistable fraction $f_B = 0$), while in the last cycle the population is almost 100% bistable (i.e. $f_B \approx 1$). The bistable population had stable steady states around the environmental optimal values (i.e. kept a similar genotype, $A^\ast(\theta) = \{20, 80\}$) regardless of the specific environmental state, while the population phenotype –epigenetically– shifts towards the current optimum after each environmental transition (Figure 4.6C). Here, the final population adapted to the fluctuating environment by finding a genotype ($\theta_B$) in the biophysical parameter space where the transition between bistable solutions was efficient through epigenetic switching.

Monostable populations can adapt genetically through mutation after a change in environment (i.e. random walk in parameter space $\theta$ to a new monostable phenotype; see Figure 3.1). If mutations are frequent and mutation step-size ($M$) is large, then genetic adaptation can occur within several generations and outcompete bistable epigenetic switching. Monostable populations have potentially lower gene expression noise cost than bistable populations (which have an increased fraction of maladapted phenotypes arising from epimutations). Thus, longer epochs in a constant environment (smaller $\nu$) might favor genetic adaptation over epigenetic switching.

To explore the transition from epigenetic switching to genetic adaptation, I simulated the dynamics of evolutionary adaptation for longer epochs (smaller $\nu$) and bigger mutation step-size (larger $M$). In these cases, the population quickly adapted to the fluctuating environment (Figure 4.7). In general, two distinct types of behaviors were observed. In some cases, the population stayed in a region of the biophysical parameter space characterized by either a low nonlinearity ($n_H \leq 1$) or a low dis-
Figure 4.6: Population dynamics in different periods of the adaptation to the fluctuating environments. Population dynamics for initial, intermediate, and final cycles in the evolutionary dynamics ($\theta_0 = \{k = 80, n_H = 1, K_D = 10\}$ with parameters $N = 10000$, $\nu = 0.1$, $s_t = 6$, $u = 0.03$, and $M = 1.1$). At each generation, I plot the joint distribution of the actual phenotype ($A$) from stochastic simulation and the expected steady state value(s) $A^*(\theta)$ from the deterministic solution given the individual’s genotype. The optimal phenotype given the environment is shown as dotted line, and the fraction of bistable individuals in the population ($f_B$) is listed in the top. The population fitness $w$ per generation for the corresponding cycles is shown on the right.

sociation constant ($K_D \approx 1$), and the adaptation after each environmental change was achieved mostly through mutations on the maximum synthesis rate $k$. If the mutation step size is large enough (e.g. $M = 5$), and the environmental fluctuations infrequent (e.g. $\nu = 0.01$), populations implementing this type of genetic adaptation can have a high $W_{cycle}$. Nevertheless, in all analyzed cases, the maximum $W_{cycle}$ was obtained when the population migrated to a high nonlinearity regime ($n_H \geq 3$; Figure 4.7B), where adaptation can occur either by epigenetic switching (i.e. bistable population; Figure 4.6), or through genetic adaptation by mutating $K_D$ (Figure 4.7).

The alternative solution regimes were also observed in the CONTROL simulations, where the gene expression is assumed deterministic and therefore the biochemical noise is excluded. In these cases, the bistable fraction $f_B$ in the population was
always low, and varied significantly over time, constantly disappearing (Figure 4.8). In other words, bistability was not selected in the absence of biochemical noise (i.e. when epimutations were impossible). Still, the $W_{\text{cycle}}$ for populations adapting in the region of the biophysical parameter space with either low $n_H$ or low $K_D$ represents only a local optimum, and the highest $W_{\text{cycle}}$ was achieved with high nonlinearity and adapting genetically through mutations on $K_D$. In general, the relative advantage of the high nonlinearity on the $W_{\text{cycle}}$ decreased as $M$ and $1/\nu$ (i.e. the epoch length) increased. For example, in Figure 4.8B only one cycle (from generation 9400 to generation 9599) had high $n_H$ and high $K_D$, with low variation in $k \approx 80$. This
Figure 4.8: Average genotype in a population adapting to the fluctuating environment in the absence of biochemical noise. The $W_{\text{cycle}}$, the fraction of bistable individuals ($f_B$), and the average biophysical parameters in the population are shown for each simulation, with initial genotype ($\theta_0$) $k = 80$, $n_H = 1$, $K_D = 10$, evolutionary parameters $N = 10000$, $s_t = 6$, and $u = 0.03$, and deterministic gene expression (i.e. CONTROL simulations). (A) $M = 2.1$ and $\nu = 0.04$. (B) $M = 5$ and $\nu = 0.01$. The population average is shown as black dotted lines, while the average on the bistable (violet) and monostable (blue) subpopulations are shown as dots for each generation. If the subpopulation was less than 1% of the population, its average genotype is excluded from the plots.

cycle had the highest $W_{\text{cycle}}$, but the benefit was small, and the genotype was quickly lost.

I verified that the high-nonlinearity genotypes were globally optimal by re-running evolutionary simulations for different initial genotypes and for more generations. In general, simulations starting with $\theta_1 = \{k = 80, n_H = 6, K_D = 45\}$ converged faster to the same global optima than simulations starting with $\theta_0$. Thus, $\theta_1$ will be used for all future simulations, unless specified otherwise.

The average biophysical parameters in each environmental state (i.e. LOW and HIGH) for different values of mutation step-size $M$ and environmental fluctuation
Figure 4.9: Average genotypes per environmental state. The color maps show the average value of the maximum synthesis rate $k$, the Hill coefficient $n_H$, and the constant $K_D$ in the population in each environment (HIGH and LOW) of 10 independent evolutionary simulations for distinct mutation step-sizes ($M$) and environmental fluctuation frequencies ($\nu$). Each simulation ran for 10,000 generations with evolutionary parameters $N = 10000$, mutation rate $u = 0.03$ and selection pressure $s = 6$. In all cases, the synthesis rate $k$ presented values around 80, with a tendency to slightly higher values in the HIGH environment, and vice versa in the LOW environment. $K_D$ showed average values between 20 and 80 in all cases, and the effect of the environmental state over the average value (i.e. the difference between the average $K_D$ in the LOW and HIGH environments) increased as $\nu$ values decreased and $M$ values increased. The nonlinearity or Hill coefficient $n_H$ appeared mostly insensitive to the environmental state, but with a strong dependency on the environmental fluctuation frequency, with higher $n_H$ values as $\nu$ decreased. The same qualitative behaviour was observed for different selection pressures (i.e. $s_T = 3$ and $s_t = 15$).
4.2.1 Monostable and bistable subpopulations

As expected, the fraction of monostable genotypes increased in simulations with higher $M$ and smaller $\nu$ (Figure 4.10A). Nevertheless, the fraction of bistable individuals in the population ($f_B$) never became zero and even showed regular temporal and environmental patterns (Figure 4.10B). This observed coexistence of subpopulations was stable. Longer simulations displayed the same pattern, and a larger population size even stabilized the observed bistable fraction $f_B$ (Figure 4.10C). In the unusual cases where a subpopulation would go extinct, it would re-appear and re-establish itself in subsequent generations (Figure 4.10D-E).

This coexistence is intriguing because these subpopulations applied different adaptation strategies. Analyzing the subpopulation dynamics, we can observe that the monostable subpopulation shifted their steady state values towards the given optimal phenotype each time the environment changed, i.e. genetic adaptation; however, the bistable subpopulation kept their expected steady state values $A^*(\theta)$ centered on the two optimal values, regardless of the current environment, and only shift their phenotype, i.e. epigenetic switching (Figure 4.11). If one strategy is more fit than the other, then we would expect one to fix and the other to go extinct. This suggests that “seeding” of one subpopulation from another through genetic mutation plays a role in the evolutionary stability of co-existing subpopulation fractions.

4.2.2 Mutational cloud effect

To further understand this phenomenon, the dynamics of the population structure, or distribution of genotypes, were also analyzed. Assuming a population in a stationary state given a fixed environment, we call the mutational cloud to the set of mutant genotypes expected to be generated every generation from the selected genotypes in the population, that being less optimal would be also excluded from the population in the next selection event. In other words, the mutational cloud is the transient
Figure 4.10: Coexistence of bistable and monostable subpopulations is an evolutionarily stable state. (A) The color maps show the average fraction of bistable individuals in the population \( \langle f_B|E \rangle_{\text{sim}} \) in each environment (HIGH and LOW) of 10 independent evolutionary simulations for mutation step-size \( M \) and environmental fluctuation frequency \( \nu \). Each simulation ran for 10,000 generations with evolutionary parameters \( N = 10,000, s_I = 6, u = 0.03, k = 80, n_H = 6, \) and \( K_D = 45 \) as the initial genotype \( \theta_1 \). (B) Density plot of \( f_B|E \) for \( \nu = 0.1 \), and \( M = 5 \) (red box in A) as a function of generations after environmental change. Each column corresponds to the distribution of the \( f_B|E \) over the entire simulation. (C) Increasing the population size to \( N = 25,000 \) sharpened the observed trends of stable co-existence. (D,E) Two examples showing how the (D) monostable and (E) bistable subpopulations can become extinct (i.e. \( f_B = 1 \) and \( f_B = 0 \), respectively; red arrows) yet are re-established in the evolutionary simulation. Both examples were for evolutionary parameters \( N = 630, s_I = 6, u = 0.03, \nu = 0.1, \) and (D) \( M = 1.7 \), (E) \( M = 5 \). I deliberately decreased \( N \) because extinction events are more common in smaller populations.
Figure 4.11: Coexistence of bistable and monostable subpopulations in the final adapted population under multiple evolutionary conditions. Population structure and evolutionary dynamics of adaptation to two environmental states (LOW, dark green; HIGH, light green) at 10,000 generations. Each simulation started from the same genotype and evolutionary parameters as Figure 4.7 with (A) $\nu = 0.04$, $M = 2.1$ (25-generation epoch, moderate mutation step-size) and (B) $\nu = 0.01$, $M = 5.0$ (100-generation epoch, large mutation step-size). The final adapted population contained both monostable and bistable subpopulations with bistable fraction $f_B$ (shown in pink). For each subpopulation (bistable, top; monostable, bottom), we plot the joint distribution of the actual phenotype ($A\ast(\theta)$) and deterministic phenotype $A\ast\theta$. As a reference, the optimal phenotypes for each generation are shown as a dotted line.

Interestingly, when coexisting subpopulations were observed, the mutational cloud associated to the selected bistable subpopulation played an essential role in the adaptation of the monostable subpopulation in the next environmental change. In these cases, every generation mutants were generated from the bistable subpopulation which lied in monostable region with a steady state optimal or close to optimal in the alternative environment. These mutants are expected, given that the selected bistable individuals have their two steady states selected to be optimal in both pos-
sible environments, and changes in the $K_D$ constant only pushes the genotype to one of these steady states (see Figure 3.4).

In order to exemplify this phenomenon, a simple experiment was performed where the Hill coefficient is fixed in the population ($n_H = 6$) and only $k$ and $K_D$ were allowed to mutate (Figure 4.12). In a population already adapted to the LOW environment, both monostable individuals with high $K_D$ values and bistable individuals were selected in the previous generation. Also, new mutant genotypes appeared in the current generation, most of them maladapted to the current environment. Particularly, many of the mutants arising from the bistable subpopulation lied in the monostable region with low $K_D$ values and, consequently, high steady state expression levels $A^*(\theta)$. If the environment would stay LOW, most of these mutants would be rejected by selection. Nevertheless, once the environment changes from LOW to HIGH, these de novo monostable individuals are already adapted and can overtake the population in a few generations. In other words, the new adapted monostable individuals resulted from the already present mutational cloud of the selected bistable subpopulation. This phenomenon was observed regardless of the direction of the environmental change (Figure 4.12).

The specific distribution of mutational cloud in the phenotypic landscape might also contribute to the asymmetry in adaptation times depending on the environmental transition type. Further analysis is required to determine if this coexistence represents simply an artifact of the genotypic and phenotypic landscape in our model or a more interesting phenomenon of adaptive variation. Nonetheless, these observations have shown that the population structure and dynamics play a fundamental role in the adaptation process, as well as that these adaptation strategies are not mutually exclusive.
Figure 4.12: Mutational cloud of the bistable subpopulation seeds the monostable subpopulation before an environmental change. A population with an initial synthesis rate $k = 80$ and $K_D = 45$, and fixed Hill coefficient $n_H = 6$ was evolved with parameters $N = 10000$, $\nu = 0.1$, $s_i = 6$, $u = 0.03$, and $M = 2$. Every generation in a constant environment, de novo mutants arise from the bistable subpopulation which lay in monostable region with an optimal or close to phenotype in the alternative environment, which can be selected once the environment changes. The genotypes selected in the previous generation are shown as filled circles: violet if the genotype is bistable, dark blue if monostable. The de novo mutants arising in the current generation are shown as empty circles, and the color corresponds to the parental genotype: pink if the mutant arose from a bistable parent, blue otherwise. The current environment is shown in the top of each graph, and the bistable region is marked by gray lines as reference. An example of a transition from LOW to HIGH environments is shown in the top row, and from HIGH to LOW in the bottom.
4.3 Discussion & Conclusions

This initial analysis has shown that under the model developed in Chapter 3, populations are capable to adapt to the environmental dynamics tested here. Both monostable and bistable genotypes were selected, with the specific proportion depending on the evolutionary conditions. Particularly, populations with only bistable genotypes over a cycle must be applying epigenetic switching as the adaptation strategy. Congruently, little temporal variation on the average maximum synthesis rate \( \langle k \rangle_g \) was observed in these populations (Figure 4.2). Moreover, the two stable steady states associated with these bistable genotypes remained around the two environmental optimal phenotypes, regardless of the current environmental state, and only the phenotypic distribution in the population adapted after each environmental transition.

Interestingly, 100% monostable subpopulations were never observed as the optimal solution to the fluctuating environment. The fraction of monostable individuals increased as the mutation step-size \( M \) and the environmental epoch length \( (1/\nu) \) increased, but a fraction of bistable individuals always coexisted. In these cases, interesting temporal and structural patterns were observed (Figure 4.10), and \textit{de novo} mutations of the bistable subpopulation might be seeding the monostable subpopulation after each environmental transition (Figure 4.12).
Characterizing adaptation strategies & their selective advantage

To better understand the adaptation process and evolutionary forces that generate co-existence, it was informative to analyze the genealogy (i.e. lineages) of the current population. By tracking the history of individual cells, I identified those lineages that persisted across multiple environmental cycles. The specific adaptation strategy implemented by these successful lineages can be characterized, and their occurrence quantified as evolutionary conditions change.

In this chapter, I start by describing how individual lineages are extracted from evolutionary simulations, as well as the signatures displayed by the distinct adaptation strategies—epigenetic switching, and genetic adaptation—in these lineages. Then I show the effect of evolutionary conditions over the dominant adaptation strategy. Finally, I analyze how a trade off on the population fitness between adaptation time and phenotypic robustness along the environmental cycles can explain the patterns on the selected strategy, and how the Hill coefficient is the dominant parameter that modulates the balance on the fitness trade-off.
5.1 Individual lineages & adaptation strategies

We chose a full environmental cycle (i.e. LOW epoch + HIGH epoch) as the minimal time unit for characterizing the lineage behavior because we are interested in adaptation to the fluctuating environment. Those individual lineages that have persisted –with or without mutations– through at least one full environmental cycle are relevant for understanding adaptation to the environmental dynamics.

5.1.1 Tracing individual’s history

I tracked the evolutionary history of the population to elucidate lineages that persisted with or without mutations over one full environmental cycle. An unfit mutant could arise at the end of a cycle or a fit genotype might go extinct due to genetic drift and gene expression noise. For instance, when analyzing the genealogical records of all cells over 2 cycles (Figure 5.1), a great diversity of genotypes was always observed in the last generation. Nevertheless, most of these genotypes would not survive the next environmental transition. The lineages and ancestral genotypes showed that the genotype diversity decreases significantly at each environmental transition, given that only a few lineages will be able to adapt (i.e. population bottlenecks). We used the behavior of the lineage on the previous full cycle (i.e. 2-cycle ancestor) to classify the individual strategy, then the lineage had been proved to survived and adapt to two environmental transitions (i.e. LOW to HIGH, as well as HIGH to LOW transitions).

If a particular adaptation strategy is successful, then we expect those lineages using that strategy to have a higher fitness and to persist over multiple environmental cycles. More than one lineage can persist over a cycle, but fewer than expected from coalescent theory because our population is evolving under selection and faces a bottleneck at each environmental transition. The weight of each persisting lineage is
Figure 5.1: **Lineage analysis of cells evolving in a fluctuating environment.** At the end of each cycle (LOW epoch + HIGH epoch), I analyzed the genealogy of cells over the past two cycles. All cells were classified based on the evolutionary strategy used by their 2-cycle ancestor over a full cycle (bigger dots). The top bars show the environmental state per epoch (dark green for LOW, light green for HIGH). I plot the number (#) of distinct lineages (solid line) and genotypes (dotted line) as a function of past generations on the top row. The middle rows plot the corresponding genotypes $\theta$ and the bottom shows the individual ancestral lineages. Ancestral genotypes can be bistable (violet) or monostable (blue). (A) Example of lineage analysis of cells that use epigenetic switching (ES) strategy for $\nu = 0.1$ and $M = 5$, i.e. their 2-cycle ancestors were fully bistable and persisted a full cycle without mutations. Note that there are distinct lineages with identical genotypes. (B) Example of lineage analysis of cells that use bistable adaptation (BA) strategy for $\nu = 0.04$ and $M = 2.1$, i.e. their 2-cycle ancestors were fully bistable but accumulated mutations over the next cycle. (C) Example of lineage analysis of cells that use genetic adaptation (GA) strategy for $\nu = 0.01$ and $M = 5$, i.e. their 2-cycle ancestors had monostable genotypes and accumulated mutations over the next cycle. In all cases, I used $N = 4000$, $s_t = 15$, and $u = 0.03$. 

124
proportional to the number of progeny at the end of the cycle.

Figures 5.2-5.4 show the population distribution of genotypes each generation for the examples displayed in Figure 5.1A-C, respectively. Only the first environmental cycle is shown in each case, and the distribution of genotypes in the surviving lineages corresponding to the same generations are also shown. In all cases, a complex dynamic can be recognized in the population structure compared to the lineages that will persist in the next environmental cycle. Three interesting phenomena can be recognized. First, the de novo mutations (i.e. mutational cloud) play an important role in seeding the transient subpopulations that generate the coexistence of monostable and bistable genotypes, as described in Section 4.2.1 (see Figures 4.10 and 4.12, and Section 4.2.2 for discussion). Second, transient subpopulations arise and increase in frequency during the first generations after an environmental transition, but are eventually lost in competition to other more optimal genotypes. Last, if the mutation step-size is large enough (e.g. $M = 5$, Figures 5.2 and 5.4), a monostable subpopulation with a high $K_D \approx 120$ value emerges in the LOW environment, but this population is lost soon after the transition to the HIGH environment because it is far from the bistable region.

5.1.2 Characterizing adaptation strategies

An epigenetic switch can adapt with no mutation; thus, lineages with a bistable genotype and no mutations during a full cycle were classified as having an epigenetic switching (ES) strategy (Figure 5.5). On the other hand, those lineages that had at least one monostable genotype and accumulated mutations between the 1-cycle and 2-cycle ancestors were classified as having a genetic adaptation (GA) strategy. Lineages with only bistable genotypes between the 1-cycle and 2-cycle ancestors that accumulated mutations during a cycle were classified as having a hybrid bistable adaptation (BA) strategy. Although some of these mutations can be neutral, most
**Figure 5.2:** Example of the population distribution of genotypes versus surviving lineages using ES strategy. The first environmental cycle from Figure 5.1A is shown. The distribution of the selected population in the genotype space with $k = 80 \pm 5$ is plotted. Scale bar is set to saturate at 10% of the population (dark blue) and one individual already appears as a yellow square. The *de novo* mutations, which appear each generation and which have not yet experienced selection, are shown as small black squares. The distribution of surviving lineages genotypes are also shown (red squares). The bistable region is delimited by a pink line. Each row corresponds to the first four generations after the transition to the LOW (dark green, top) and HIGH environment (light green, bottom), as well as the last two generations in each epoch.

**Figure 5.3:** Example of the population distribution of genotypes versus surviving lineages using BA strategy. The first environmental cycle from Figure 5.1B is shown using the same notation as Figure 5.2.
5.1.3 Evolutionary conditions & the selected adaptation strategy

Lineage analysis demonstrated that distinct strategies were favored in different conditions (Figure 5.6). If the environment fluctuated frequently (i.e. high $\nu$ values), the dominant adaptation strategy was ES. In slowly fluctuating environments (i.e. low $\nu$ values), GA was used if the mutation step-size ($M$) was large enough; otherwise, BA was the dominant adaptation strategy.

Interestingly, I observed a mixture of strategies across lineages (no strategy represented 100% of cases), and the transition between preferred adaptation strategies as a function of $\nu$ and $M$ was gradual. As expected, the simulations in these regimes showed a high temporal variation on the fraction of adaptation strategies in the parental cycles (Figure 5.7).

Our CONTROL simulations with deterministic dynamics (where no stochastic epigenetic switching can occur even if the system is bistable) showed that none of
**Figure 5.5: Adaptation strategies.** Illustration of epigenetic switching (ES), bistable adaptation (BA), and genetic adaptation (GA) strategies and underlying genotypes with fixed synthesis rate ($k = 80$). The bistable region of the genotype space is highlighted in pink. The phenotype distribution $\rho(A)$ for each genotype ($\theta$) is shown in the inset, both in linear and logarithmic scale. BA is bistable, as seen in logarithmic scale, but appears effectively monostable in linear scale. This arises because $K_D$ evolves each epoch to favor one mode over the other by decreasing the relative rates of epigenetic switching between the largest and smallest mode.

The observed trends in Figure 5.6 persisted over a range of different evolutionary parameters, although the location of the boundaries between the dominant strategy would shift. For example, increasing the selection pressure ($s_t$) or the mutation rate ($u$) shifted boundaries to favor GA, whereas increasing population size ($N$) favored ES (Figure 5.8).

As discussed in the previous chapter (see Section 4.2.1), the average fraction of bistable genotypes $\langle f_B \rangle_{\text{sim}}$ never decreased to zero as the evolutionary parameters varied, even when GA was the preferred lineage strategy (e.g. low $\nu$, high $M$; Figure
Figure 5.6: Each adaptation strategy is favored under different evolutionary conditions and the transition between selected strategies is gradual. In the left, each color map shows the fraction of parental lineages using a specific adaptation strategy (epigenetic switching, ES; bistable adaptation, BA; or genetic adaptation, GA) averaged over all cycles and ten independent replica simulations for the corresponding mutation step-size ($M$) and environmental fluctuation frequency ($\nu$). Each simulation ran for 10000 generations with evolutionary parameters $N = 10000$, $s_t = 6$, $u = 0.03$ and $k = 80$, $n_H = 6$ and $K_D = 45$ as the initial genotype ($\theta_1$). This initial genotype sped up evolutionary simulations by being closer to final selected genotypes in all simulations. In the middle column, the equivalent results of the CONTROL simulations, where gene expression dynamics are deterministic and no stochastic epigenetic switching can occur, are shown. All lineages exhibited GA and neither bistable strategy (ES or BA) was ever selected. The corresponding bistable fraction ($\langle f_B \rangle_{\text{sim}}$) averaged over all cycles and ten independent replica simulations for the stochastic simulations (top) and deterministic CONTROL (bottom) are shown in the right.
Figure 5.7: Different adaptation strategies co-exist and exhibit large fluctuations over evolutionary time. Each plot shows the fraction of parental lineages with different adaptation strategies (epigenetic switching, ES; bistable adaptation, BA; genetic adaptation, GA) per environmental cycle for mutation step-size ($M$) and environmental fluctuation frequency ($\nu$). In each case, only the last 50 cycles are shown. Each simulation ran for 10,000 generations with evolutionary parameters $N = 10000$, $s_t = 6$, $u = 0.03$, and $k = 80$, $n_H = 6$, and $K_D = 45$ as the initial genotype ($\theta_1$).

These results arise from the increased seeding of genetic mutants (which is facilitated by higher $M$) from the monostable to bistable subpopulation (see Section 4.2.2). Conversely, more neutral or near neutral mutations are expected in the bistable region for smaller mutational step-sizes ($M$). It is exactly in this regime where the fraction of ES parental lineages decreased whereas BA increased. However, most lineages displaying BA in this regime also persisted into the next cycle without accumulating any mutations and, thus, automatically switched to ES.

To determine the transition rate between evolutionary strategies, I analyzed lineages back to the 3-cycle ancestor. I calculated the statistics of transitions between adaptation strategies by comparing strategies in 1-cycle and 2-cycle ancestors (cur-
Figure 5.8: Increasing selection pressure or mutation rate favors genetic adaptation, whereas increasing population size favors epigenetic switching. Each color map shows the average fraction of parental lineages using each adaptation strategy (epigenetic switching, ES; bistable adaptation, BA; genetic adaptation, GA) for the same range of mutation step-size ($M$) and environmental fluctuation frequency ($\nu$) as Figure 5.6. Evolutionary parameters used in Figure 5.6 ($s_t = 6$, $N = 10000$, $u = 0.03$) are highlighted in red boxes. (A) The effect of only changing the selection pressure ($s_t$) over three evolutionary replicas. (B) The effect of only changing the population size ($N$) over three evolutionary replicas. (C) The effect of only changing the mutation rate ($u$) over ten evolutionary replicas. All simulations ran for 10000 generations with $k = 80$, $n_H = 6$, $K_D = 45$ as the initial genotype ($\theta_1$).
rent adaptation strategy) versus 2-cycle and 3-cycle ancestors (previous adaptation strategy) and averaged over all cycles (Figure 5.9). Only one strategy appeared with a high probability of displaying the same strategy over consecutive cycles, suggesting one strategy is dominant for each evolutionary condition. In cases where a significant fraction of alternative strategies was observed, the majority of the lineages switched back to the “dominant” adaptation strategy in the following cycle. For example, in the case with small mutation step-size ($M = 1.1$) and large environmental fluctuation frequency ($\nu = 0.1$), from all the lineages displaying BA in the previous cycle, $\approx 61\%$ persist the following cycle using ES. On the other hand, decreasing $\nu$ (0.04), BA becomes the dominant strategy and $\approx 67\%$ of the lineages displaying ES in the previous cycle switched back to BA.

I tested the robustness of our results to alternative choices and assumptions in the evolutionary model (see Section 3.4). In general, the same qualitative behavior was observed (Figure 5.10). For instance, different selection schemes were implemented obtaining the same qualitative results: Truncation selection, where only a certain fraction of the best individuals can be selected, each with the same probability, resulted in practically identical results. Proportional and weighted selection, where the probability of an individual to be selected is simply proportional to its fitness value, reduced the ES region. Noteworthy, the selection pressure cannot be tuned in these two selection schemes, and the results were similar to the tournament selection with higher selection pressure (Figure 5.8A). Importantly, allowing the environment to fluctuate randomly between the two possible states with mean frequency $\nu$ had more variable results, but no significant change was observed in the qualitative trend. Using a Moran model format, where the reproduction and death events are treated as stochastic events instead of assuming non-overlapping generations (Wright-Fisher format) the same qualitative behavior was observed in the adaptation strategies selected. Noteworthy, these simulations are considerably
Figure 5.9: Transitions between adaptation strategies as a function of evolutionary parameters. (A) An example showing how the current and previous adaptation strategies were defined for a surviving population. In this particular simulation, the population displayed multiple strategies in the ancestral lineage between the 2-cycle and 1-cycle ancestors (current adaptation strategy): 63.825% epigenetic switching (ES), 35.95% bistable adaptation (BA), and 0.225% genetic adaptation (GA). However, the ancestral lineage between the 3-cycle and 2-cycle ancestors (previous adaptation strategy) were 100% ES. (B) The color maps show the percentage of ancestral lineages that displayed one adaptation strategy (current adaptation strategy) and other adaptation strategy in the preceding ancestral lineage (previous adaptation strategy). These statistics were calculated for 10 evolutionary replicas for mutation step-size ($M$) and environmental fluctuation frequency ($\nu$). Each simulation was run 10,000 generations with evolutionary parameters $N = 10000$, $s_t = 6$, $u = 0.03$, and $k = 80$, $n_H = 6$, and $K_D = 45$ as the initial genotype ($\theta_1$).

more expensive computationally, so a smaller population size and shorter simulation times were used. Similarly, the same behavior was obtained by changing the fitness function to a Gaussian or a step-like function with similar span around the optimal phenotypes, or using the average protein number or the distribution of protein numbers over the individual lifespan as phenotype. Substituting the used spherically symmetric 3D mutation scheme for a homogeneous spherical mutation or a
homogeneous cubic mutation scheme produced the same qualitative results.

Moreover, even if quantitative aspects of the model changed the value of evolutionary parameters values where epigenetic switches were selected, the observed shifts were congruent with the expectations of our model (Figure 5.10). For example, the rate of epimutation is sensitive to the frequency and magnitude of stochastic events. The magnitude of stochastic events is inversely proportional to the total number of molecules. Thus, we expect a higher rate of epimutation for smaller numbers of molecules. The rate of epimutation should also increase as the two modes become closer. Then, either choosing optimal phenotypes with smaller values (e.g. \( A_{\text{opt}} = [10, 40] \)), or increasing the basal activity value (e.g. \( \alpha = 0.3 \))—adjusting the optimal phenotypes accordingly to still have possible bistable solutions (\( A_{\text{opt}}^{(L)} = \alpha \cdot A_{\text{opt}}^{(H)} \); see below)—favor epigenetic switching, increasing the region of the color map where this strategy is dominant. The opposite effect is observed if the \( A_{\text{opt}} \) values increase (e.g. \([40, 160]\)) or \( \alpha \) decreases (e.g. 0.20). Finally, the protein degradation rate (\( \gamma \)) sets the timescale between stochastic events (i.e. faster protein degradation leads to more stochastic events per unit time during a Gillespie simulation). Thus, we expect a higher rate of epimutation for larger \( \gamma \). In all tested cases, a higher rate of epimutation favored ES over GA.

At high levels of nonlinearity, the lowest protein level is \( k \cdot \alpha \) and the highest protein level is \( k \). A bistable, epigenetic switch has two solutions, each well-adapted to one of the environments only when the ratio \( R = A_{\text{opt}}^{(L)}/A_{\text{opt}}^{(H)} = \alpha \) (Figure 3.4). Any mismatch between \( \alpha \) and \( R \) will disfavor epigenetic switching because an epimutation from an adapted mode will jump to a maladapted mode, after which the descendants must accumulate genetic mutations to adapt further. In all the main simulations presented here, \( R = \alpha = 0.25 \) was fixed. I verified that by allowing the basal activity parameter \( \alpha \) to evolve during the adaptation process, I obtained the same qualitative results (Figure 5.11). As expected, in the cases where bistability was selected —i.e.
Figure 5.10: The same qualitative trends on the selection of adaptation strategies per evolutionary condition is maintained in a wide variety of alternative model assumptions. Each color map shows the population average fraction of parental lineages using each adaptation strategy (epigenetic switching, ES; bistable adaptation, BA; genetic adaptation, GA) for the same range of mutation step-size ($M$) and environmental fluctuation frequency ($\nu$) as Figure 5.6. Differences in assumptions or parameters are listed above each plot. All values are the average of 3 evolutionary replicas of simulations run 10,000 generations with $N = 4,000$, $s_t = 6$, $u = 0.03$ and $k = 80$, $n_H = 6$, and $K_D = 45$ as the initial genotype ($\theta_1$). The exceptions are the weighted and proportional selection schemes where the selection pressure ($s_t$) cannot be tuned, and the Moran model case which ran for 1,000 generations. When the basal activity ($\alpha$) was changed, I adjusted the low optimal phenotype such that the ratio of $A^{(L)} = \alpha \cdot A^{(H)}$, where $A^{(H)} = 80$. See Section 3.4 for an explicit description of assumptions and parameters.
epigenetic switching or bistable adaptation strategies, $\alpha$ evolved to values very close to $R = A_{opt}^{(L)}/A_{opt}^{(H)} = 0.25$.

**Figure 5.11:** Allowing basal activity ($\alpha$) to evolve does not qualitatively change our results. (A) The color maps show the average fraction of parental lineages using each adaptation strategy (epigenetic switching, ES; bistable adaptation, BA; genetic adaptation, GA) of 3 evolutionary replicas under the corresponding mutation step-size ($M$) and environmental fluctuation frequency ($\nu$). Each simulation ran for 10000 generations with evolutionary parameters $N = 10000$, $s_t = 6$, $u = 0.03$ and $k = 80$, $n_H = 6$, $K_D = 45$ and $\alpha = 0.25$ as the initial genotype ($\theta_1$). The corresponding (B) average bistable fraction ($\langle f_B \rangle_{\text{sim}}$) and (C) average $\alpha$ ($\langle \alpha \rangle_{\text{sim}}$). For some examples, the dynamics over time for the geometric mean fitness per cycle ($W_{\text{cycle}}$), the average basal activity ($\langle \alpha \rangle_{\text{cycle}}$), and the average bistable fraction ($\langle f_B \rangle_{\text{cycle}}$) per cycle, as well as the fraction of parental lineages using ES as the adaptation strategy per cycle, are shown: (D) $\nu = 0.1$ and $M = 5$; (E) $\nu = 0.02$ and $M = 2.6$; (F) $\nu = 0.002$ and $M = 5$; and (G) $\nu = 0.01$ and $M = 5$.

### 5.2 Fitness costs: Adaptation time vs Phenotypic robustness

My simulations showed that different portfolios of strategies are favored for different mutation step sizes ($M$) and environmental transition frequencies ($\nu$). Bistable
strategies (ES or BA) required genotypes with large nonlinearity ($n_H$), whereas GA genotypes are presumably less constrained. However, all evolutionary simulations (including GA) evolved to average genotypes with high nonlinearity $\langle n_H\rangle_{\text{sim}}$ (Figure 5.12A). High nonlinearity was also selected in the deterministic CONTROL simulations, where ES and BA cannot occur (Figure 5.13A). The evolution to genotypes with high nonlinearity is consistent with previous work (Kuwahara and Soyer, 2012) (Section 5.2.1). Strikingly, $\langle n_H\rangle_{\text{sim}}$ was mostly determined by the environmental fluctuation frequency and it increased as $\nu$ decreased (i.e. longer epochs; Figure 5.12A). This same trend was observed when I measured the geometric mean fitness averaged over the whole simulation $\langle W_{\text{cycle}}\rangle_{\text{sim}}$ (Figure 5.12B). Populations that evolved in slowly fluctuating environments had higher average fitness and higher nonlinearity.

The geometric mean fitness per cycle, which includes the lower fitness of the transient, maladapted population immediately after a transition in addition to the higher fitness of the adapted population, was surprisingly insensitive to the mutation step-size ($M$). One might expect that a larger mutation step size increases the mutational load and, thus, decreases the average fitness of the final adapted population. This discrepancy suggested to us that the population fitness cost of new mutations (mutational load) was minimal for our mutation rate ($u$) when compared to other costs (e.g. gene expression noise and/or epimutational load). In agreement with this hypothesis, increasing the mutation rate ($u$) generated $M$-dependence in fitness, whereas decreasing the mutation rate had little effect (Figure 5.14). Varying the selection pressure ($s_t$) only affected the actual $\langle W_{\text{cycle}}\rangle_{\text{sim}}$ values observed, but not the trend (Figure 5.15). Similarly, varying the population size ($N$) had no significant effect on the population fitness behavior, except that results became noisier as $N$ decreased (Figure 5.16).

The trend in average geometric mean fitness $\langle W_{\text{cycle}}\rangle_{\text{sim}}$ as a function of $\nu$ can
Figure 5.12: Population fitness depends on environmental fluctuation frequency and reflects a trade-off between adaptation time and phenotypic robustness. (A) Hill coefficient \( \langle n_H \rangle_{\text{sim}} \) averaged over ten independent replica simulations for mutation step-size \( (M) \) and environmental fluctuation frequency \( (\nu) \). (B) Average geometric mean fitness per cycle \( \langle W_{\text{cycle}} \rangle_{\text{sim}} \) for the same simulations. (C) Average population fitness at the second generation \( \langle w \rangle_{+2g} \) and (D) ninth generation \( \langle w \rangle_{+9g} \) after an environmental transition. Evolutionary strategies with a faster adaptation time had larger fitness after the transition (C) whereas those with robust phenotypes tended to have a larger fitness once they adapted to the new environment (D). All simulations ran for 10,000 generations with identical evolutionary parameters and initial genotypes as in Figure 5.6.

be understood by considering the timescales of epochs. The population fitness is always low and maladapted when the environment first changes. Selection favors those genotypes which produce phenotypes that better match the selection pressure in the new environment. In this first phase (the “adaptation phase”), the fitness increases as the population adapts through genetic adaptation (GA and BA, mutation) and/or epigenetic switching (BA or ES, epimutation). The average population fitness two generations after each environmental change \( \langle w \rangle_{+2g} \) is \( M \)-independent and highest for the ES genotypes that emerge for large \( \nu \) (Figure 5.6B and Figure 5.12C). The rate of epimutation is faster than genetic mutation and, thus, ES (the
dominant strategy) adapts more quickly and has a higher transient fitness. However, all populations eventually reach a higher fitness and become adapted to the new environment. In this second phase (the “constant phase”), purifying selection maintains the optimal phenotype against perturbations from gene expression noise, mutations, and/or epimutation. The average population fitness nine generations after each environmental change \( \langle w \rangle_{+9g} \) is mostly \( M \)-independent and highest for small \( \nu \) (Figure 5.12D). The population spends proportionally more time in the constant phase as \( \nu \) decreases, which favors those strategies with robust phenotypes (i.e. minimize frequent, maladapted phenotypes that arise from epigenetic switching). Thus, \( \langle W_{cycle} \rangle_{sim} \) increases and GA or BA genotypes emerge as \( \nu \) decreases because the population spends more time in the “constant phase” and epigenetic switching does
Figure 5.14: Increasing the mutation rate makes population fitness sensitive to mutation step-size \( (M) \) and affects qualitative trends in adaptation time and phenotypic robustness. (A,E) Average population Hill coefficient \( \langle n_H \rangle_{\text{sim}} \) of 3 evolutionary replicas for mutation step-size \( (M) \) and environmental fluctuation frequency \( (v) \). (B,F) Average geometric mean fitness per cycle \( \langle W_{\text{cycle}} \rangle_{\text{sim}} \) for the same simulations. (C,G) Average population fitness at the second generation \( \langle w \rangle_{+2g} \) and (D,H) ninth generation \( \langle w \rangle_{+9g} \) after an environmental transition. All simulations ran for 10,000 generations with identical evolutionary parameters and initial genotypes as in Figures 5.6 and 5.12, except for mutation rates \( (u) \) listed above each plot.

not occur or is effectively zero (see \( \rho(A) \) in Figure 5.5A).

These results suggest a trade-off in the evolutionary process between minimizing the adaptation time during the adaptation phase and increasing the robustness of the phenotype during the constant phase (Figure 5.12). This trend correlates with the selected Hill coefficient value in the population \( \langle n_H \rangle_{\text{sim}} \), which suggests that natural selection tunes this trade-off via this biophysical parameter (Figure 5.17). The trade-off between two fitness costs also explains the observed gradual transition between selected adaptation strategies (Figure 5.6B). These fitness costs are continuous and can have very similar values, such that genetic drift will dominate during the selection process. As expected, the simulations in regimes with co-dominant strategies showed a high temporal variation in the fraction of adaptation strategies each evolutionary cycle (Figure 5.7).
Figure 5.15: Increasing the selection pressure increases the population fitness without affecting the qualitative trends in adaptation time and phenotypic robustness. (A,E) Average population Hill coefficient \( \langle n_H \rangle_{\text{sim}} \) of 3 evolutionary replicas for mutation step-size (\( M \)) and environmental fluctuation frequency (\( \nu \)). (B,F) Average geometric mean fitness per cycle \( \langle W_{\text{cycle}} \rangle_{\text{sim}} \) for the same simulations. (C,G) Average population fitness at the second generation \( \langle w \rangle_{+2g} \) and (D,H) ninth generation \( \langle w \rangle_{+9g} \) after an environmental transition. All simulations ran for 10,000 generations with identical evolutionary parameters and initial genotypes as in Figures 5.6 and 5.12, except for selection pressure (\( s_t \)) listed above each plot.

5.2.1 Effect of Hill coefficient

Starting from an arbitrary initial condition and with different evolutionary parameters, all populations evolved to a similar region of genotype space with higher fitness (Section 4.1). These final genotypes often had \( k \sim 80, 20 < K_D < 80 \), and large \( n_H \) (Figure 4.9). To better understand the forces that select for large nonlinearity, I considered a simple model of an infinite, clonal population with steady-state protein distribution \( \rho(A) \) determined by \( n_H, K_D \) and fixed \( k = 80 \). I calculated the expected fitness in LOW and HIGH environment as a function of \( n_H, K_D \) (Figure 5.17). My simple model demonstrates that the fitness of genotypes \( \theta \), with steady state phenotypes \( A^*(\theta) \) close to the optimal values, always increases with \( n_H \) regardless of \( K_D \), the specific environmental state (LOW or HIGH), or whether the genotype is bistable.
Figure 5.16: Decreasing the population size increases the variation between independent replicas, without affecting the qualitative trends in fitness, adaptation time and phenotypic robustness. (A,E) Average population Hill coefficient $\langle n_H \rangle_{\text{sim}}$ of 3 evolutionary replicas for mutation step-size ($M$) and environmental fluctuation frequency ($\nu$). (B,F) Average geometric mean fitness per cycle $\langle W_{\text{cycle}} \rangle_{\text{sim}}$ for the same simulations. (C,G) Average population fitness at the second generation $\langle w \rangle_{+2g}$ and (D,H) ninth generation $\langle w \rangle_{+9g}$ after an environmental transition. All simulations ran for 10,000 generations with identical evolutionary parameters and initial genotypes as in Figures 5.6 and 5.12, except for population size ($N$) listed above each plot.

The increased fitness arises from better buffering of gene expression (output) against intrinsic biochemical noise in the protein levels (input). As the nonlinearity increases, the gene expression rate $f(A)$ both above and below $K_D$ becomes more zero-order and less sensitive to fluctuations in protein levels ($A$). For the same reason, mutations in $K_D$ are less likely to affect steady-state protein levels and will have a smaller effect on fitness. Thus, increasing nonlinearity leads to higher fitness because of increased robustness of gene expression to both biochemical noise and some de novo mutations.

Analogously, increasing the Hill coefficient decreases the rate of epimutations in bistable systems (Figure 5.18). As discussed in Section 3.3.1, once in the bistable region, increasing the Hill coefficient $n_H$ can have a significant effect on the stationary distribution of the protein number ($P^*(A)$), even if the protein number associated
**Figure 5.17:** Increasing $n_H$ increases the average population fitness in both environments. (A) Contour plots as a function of biophysical parameters with fixed $k = 80$ where the steady state ($A^*$) for LOW are $A^{(L)} \pm 1\%$ (dark green) and for HIGH are $A^{(H)} \pm 1\%$ (light green). (B) I analytically calculated the average fitness of an infinite, clonal population with stationary protein distribution $\rho(A)$ given $n_H$, $K_D$ and fixed $k = 80$. The normalized distribution $\rho(A)$ was estimated numerically for each set of biophysical parameters (see Section 3.3.1). I calculated the expected fitness in each environment by integration, $E(w(E)) = \sum_{a} w^{(E)}(a) \cdot \rho(a)$.

with the modes and the deterministic steady state values ($A^*(\theta)$) stay mostly undisturbed (Figure 5.18A). For instance, the relative proportion of the modes, i.e. the probability of being each “epiphenotype”, is directly affected by $n_H$. Moreover, the rate of stochastic transition from state to the other clearly decreases as $n_H$ increases, regardless of the direction of the transition (i.e. LOW to HIGH, and HIGH to LOW phenotypic states; Figure 5.18B).

### 5.3 Discussion & Conclusions

Previous theoretical work established that optimal long-term growth occurs when the phenotype switching rate matches the environmental switching rate (Lachmann and Jablonka, 1996; Thattai and Oudenaarden, 2004; Kussell et al., 2005; Kussell and Leibler, 2005). The phenotype could switch either due to genetic adaptation with a rate that depends on the mutation rate and mutation step-size, or due to epigenetic switching (epimutation) with a rate determined by the underlying molecular system.
**Figure 5.18:** Increasing $n_H$ decreases the rate of epimutations. (A) Effect of the Hill coefficient $n_H$ value over the stationary distribution of protein number ($\rho(A)$) with $k = 80$ and $K_D = 45$ fixed. The monostable cases are shown as dotted lines. (B) The stochastic switching rate from the LOW to HIGH phenotypic states ($1/\tau_{L\rightarrow H}$, light green) and from HIGH to LOW phenotypic states ($1/\tau_{H\rightarrow L}$, dark green) as the Hill coefficient $n_H$ increases. Only the bistable cases are shown (i.e. $n_H \geq 4$).

In the natural world, epimutation rates are often faster than genetic mutation rates (Rando and Verstrepen, 2007), which suggests that fast fluctuating environments might select for epigenetic switching (ES) over genetic adaptation (GA). Previous models did not integrate and/or evaluate these two competing processes in a population of cells evolving in a fluctuating environment.

To this end, Soyer and colleagues first used computer simulations to evolve a population of self-reinforcing gene circuits (which can exhibit epigenetic switching and genetic adaptation) in a fluctuating environment (Kuwahara and Soyer, 2012; Steinacher et al., 2016). Their work demonstrated the emergence of ES (bistable genotypes) in fast fluctuating environments. However, the authors proposed that ES was accidental and that bistability emerged as a byproduct of selection for increased nonlinearity and higher evolvability (i.e. large changes in phenotype with small changes in genotype).

My simulations confirm that ES emerges in fast fluctuating environments across many different conditions (Figure 5.6). In agreement with Soyer and colleagues,
our populations also evolved to genotypes with high nonlinearity (Figure 5.12A). However, I disagree that ES is an accidental byproduct of selection for increased nonlinearity and evolvability. Bistable genotypes were only favored in the presence of gene expression noise when cells can stochastically switch phenotypes (i.e. non-zero epimutation rate). In CONTROL simulations (i.e. deterministic dynamics with no noise and, hence, no epimutation), I only observed GA (Figure 5.6C). I conclude that ES is selected over GA in fast fluctuating environments precisely because of the benefits of epimutation.

The average population fitness was mostly determined by the environmental fluctuation frequency ($\nu$), where fitness decreased as the epoch length decreased (Figure 5.12). This agrees with previous theoretical work, which showed that the evolutionary dynamics are governed by environmental dynamics (Lachmann and Jablonka, 1996; Carja et al., 2014b). In my simulations, the correlation between fitness and epoch length arises from the balance between two competing fitness costs: reducing the time required to adapt every time the environment changes (adaptation phase) and increasing the phenotypic robustness when the environment is fixed (constant phase). The ES strategy is selected at high $\nu$ (when the population spends proportionally more time in the adaptation phase) because it has a faster adaptation time (Figure 5.12C) at the cost of lower phenotypic robustness (Figure 5.12D). My work suggests that the trade-off between adaptation time and phenotypic robustness is mostly modulated by the nonlinearity. A smaller $n_H$ increases the stochastic transition rates of ES at high $\nu$ and a larger $n_H$ increases the phenotypic robustness of GA at low $\nu$.

As the environmental fluctuation frequency decreased, the selective advantage of ES disappeared and other adaptation strategies emerged. GA was used by the population when the mutation step-size was large (high $M$) and the environment fluctuated slowly (low $\nu$). My evolutionary simulations, which include population
dynamics, mutation-selection, and a continuous spectrum of genetic mutations, revealed that ES and GA need not be mutually exclusive. A hybrid strategy (BA) applying both GA and ES was observed for small mutation step-size (low \( M \)) and slow environmental fluctuations (low \( \nu \)). Abstract models that only consider discrete fixed phenotypes with stochastic transitions overlook the potential of hybrid strategies that might occur in real epigenetic systems (Kussell et al., 2005; Salathé et al., 2009; Visco et al., 2010; Liberman et al., 2011; Libby and Rainey, 2011; Carja and Feldman, 2012; Carja et al., 2013, 2014b,a; Lin et al., 2015; Belete and Balázsi, 2015).

All evolved bistable and monostable genotypes were relatively close to each other (Figure 5.5A). My simulations had a relatively high mutation rate (\( u \)) such that bistable genotypes could mutate to monostable genotypes, and vice versa. The elevated rate of seeding between these subpopulations made it challenging to distinguish whether ES (bistable) was being selected for. It has been shown previously that individual history or genealogy can efficiently reveal hidden selection forces. For example, Kussell and colleagues (Lin et al., 2015; Lambert and Kussell, 2015) demonstrated that selective pressures on a population, such as those imposed by a fluctuating environment, can be efficiently quantified by measurements on the surviving lineages. More recently, (Cerulus et al., 2016) used life-history traits of cellular growth to show that high single-cell variance in growth rate can be beneficial for the population, and that this benefit depends on the epigenetic inheritance of the growth rate between mother and daughter cells.

To this end, I analyzed the strategy of lineages across multiple cycles during my simulations. Lineage analysis demonstrated that apparent coexistence of bistable and monostable subpopulations was a transient phenomenon, and one type of strategy was typically dominant across lineages (Figure 5.6). My analysis suggests that population snapshots (e.g. bimodal versus unimodal distribution of phenotypes) can
miss the contribution of epigenetic switching. Future experimental studies on the evolution of epigenetic switches might consider analyzing lineages using time-lapse microscopy, as done by Balaban et al. (2004).

5.4 Contributions

Chapters 3, 4 and 5 were modified from a manuscript by Mariana Gomez-Schiavon and Nicolas E. Buchler, which is currently under review for publication (e-print available at bioRxiv, doi: http://dx.doi.org/10.1101/072199). The model design and mathematical analysis were done by Mariana Gomez-Schiavon. Both authors wrote the manuscript. This work was funded by a CONACYT graduate fellowship (MGS), the National Institutes of Health Directors New Innovator Award DP2 OD008654-01 (NEB), and the Burroughs Wellcome Fund CASI Award BWF 1005769.01 (NEB).
6

Future directions

6.1 Stochastic dynamics of early transcriptional response

6.1.1 A more complex model of gene expression

The fluorescence intensity of active transcription sites \textit{Fos} exhibited significant variability (Dr. Anne West, personal communication). This suggests that the IEG promoters might display a more complex activity profile than just the ON and OFF discrete states. Changes in histone acetylation are known to take part on the activation of IEGs (West and Greenberg, 2011). Recent work has shown that chromatin regulators induced discrete expression states in the regulated genes (Bintu et al., 2016), and in some cases three distinct states were detected. Moreover, a third expression state has been proposed for the \textit{Fos} gene in other human tissue (human osteosarcoma cells) (Senecal et al., 2014). Our method can be easily extended to consider a third promoter state, and this new model can be compared with the simpler models using the information criterion metrics described in Section 2.3.3. However, the state reaction matrix associated with a 3-states model is 4 times larger than the 2-states model. Our current Bayesian inference algorithm with 2-states model is slow. Thus, a more efficient algorithm (see below) will be required to efficiently...
explore the 3-states promoter hypothesis.

A series of molecular processes occur before a mature mRNA is released to the cytoplasm, where it can be detected as a single fluorescent dot (see Section 2.2.3). For example, RNA polymerase moves at $34 \pm 11\text{bp/s}$ (Bahar Halpern et al., 2015) and transcription of Fos (3461 bases) should take 100 secs. Additionally, the mRNA needs to be released from the RNA polymerase complex, processed (e.g. removing intron sequences), and exported from the nucleus to the cytoplasm. Thus, there is a several minute delay between the start of transcription and appearance of a final, cytoplasmic mRNA. Given that we are interested in early time dynamics (e.g. 5 minutes after stimulus), the time delay between promoter state (i.e. initiation of transcription) and cytoplasmic mRNA could have a significant effect on distribution of mRNAs and, thus, the estimated biophysical parameters. The delay between the initiation of transcription and the detection of mature mRNA with smFISH might be incorporated into our gene expression model. This delay can be incorporated in the CME (i.e. delay CME, or DCME). Recently, some tools have been developed to solve DCME which can be applied to our gene expression model (Leier and Marquez-Lago, 2015). Another possibility is to include an additional species into the model representing the nascent mRNA with a certain rate of “transition” to a mature state. Nevertheless, an additional species will increase the number of possible states in the system significantly, increasing the size of the state reaction matrix, and slowing down the Bayesian inference algorithm. Once again, a more efficient algorithm is required.

6.1.2 A more efficient algorithm for Bayesian inference

As discussed at the end of Chapter 2, the computationally expensive step during Bayesian inference is the calculation of the matrix exponential, which is required each iteration to obtain the probability distributions for the gene expression after
stimulus for a proposed set of kinetic parameters. As proposed in Section 2.7, an *ad hoc* algorithm might be implemented, taking advantage of the sparsity of the state reaction matrix and the fixed set of nonzero components.

The Metropolis Random Walk (MRW) algorithm could be further optimized. For instance, the covariance matrix $\Sigma$ used to draw the proposal parameter set each iteration (see Section 2.3.2) can be fine-tuned for efficient sampling. The values of $\Sigma$ determine the ratio between the acceptance and rejection of proposals. If $\Sigma$ is too large, most of the proposals will result in a rejection, and more iterations will be need to accurately estimate the posterior distribution. Analogously, if $\Sigma$ is too small, more iterations will be required to sufficiently explore the parameter space. An optimal value of $\Sigma$ will minimize the run-time of the MRW (MacKay, 2003). This value will depend of the specific model and experimental data being analyzed, and $\Sigma$ has to be determined for each specific case. Adaptive Monte Carlo methods (e.g. adaptive Metropolis), where the variance in the proposal (e.g. $\Sigma$) is optimize during the algorithm run, can be implemented. Nevertheless, adaptive algorithms can introduce biases, and a careful implementation and validation is required.

The Metropolis method is a simple but flexible approach, and for this reason was chosen for our model. Nevertheless, multiple Monte Carlo methods exist, each with different strengths and limitations. For example, *Gibbs sampling* uses conditional probabilities from multiple blocks or dimensions of the system’s space to draw the proposals. Gibbs algorithm can speed up convergence. Nevertheless, this algorithm requires to be able to sample from the conditional probability functions, which might be hard to define in our system. Alternatively, *slice sampling* method also divides the state space in blocks or dimensions, but the conditional probabilities are not required; instead, a similar approach to MRW is used for generating proposals, but the variance or step-size “self-tunes” during the algorithm run. Furthermore, these two algorithms, Gibbs and slice sampling, also have the advantage that by changing
dimensions independently (which can be arbitrarily defined), the matrix exponential might not have to be calculated at each iteration. The hybrid—or Hamiltonian—Monte Carlo method uses the gradient around the current state to select the proposal, potentially converging faster to the target distribution. But the gradient needs to be evaluated, which might be complicated in a complex multi-dimensional system, as the one presented here. Finally, the Skilling’s multi-state leapfrog method also takes advantage of a continuous space, but without requiring the gradient. This algorithm maintains a set of states, and each proposal is a linear transformation of these states. The Skilling’s method can help to accelerate convergence for systems with strong correlated parameters (MacKay, 2003).

The implementation of a graphical model should also be investigated. Graphical models allow to factorize high-dimensional systems into a series of conditional distributions, then simplifying the sampling process and the estimation of the posterior distribution. In particular, applied to our model, it might help reducing the number of times the matrix exponential needs to be calculated by separating the optimization process of the parameters involved in the state reaction matrix of induced conditions from the rest of parameters.

The method presented in Chapter 2 is flexible enough to be applied to a wide variety of gene expression systems, always that the single cell measurement of active transcription sites and mRNA molecules is available. Creating a stable, portable and user friendly interface will facilitate the use of our algorithm by other experimentalist research groups. For instance, I am currently developing a library and running instructions in c++ to implement the algorithm presented in Chapter 2. The idea is that the user will only require to install the library, compile the program, define the input parameters (e.g. model, initial parameters, number of iterations, and covariance matrix), and a file with the MRW results will be produced. Posterior improvements might include the development of a cross-platform graphical applica-
tion (i.e. GUI) to facilitate the use of our algorithm by experimentalists. The Qt c++ framework can be used to build a GUI associated to the previously developed c++ library.

6.1.3 Using prior biological knowledge to constrain Bayesian inference

For simplicity, we assumed that all prior probabilities $P(\theta|N)$ were equal for all parameter sets $\theta$ (see Section 2.3.2). Nevertheless, we can set priors to enforce certain biological knowledge. For example, a higher mRNA synthesis rate is expected after stimulus compared to the uninduced condition. Thus, the priors can be set to penalize parameter sets where this relation (i.e. $\mu > \mu_0$) is not satisfied. Additionally, if specific perturbations are performed to the experimental system (e.g. addition of a drug with a known effect), the expected behavior can be incorporated to the parameter estimation through these prior probabilities.

6.2 Theoretical framework for the evolution of epigenetic switches in fluctuating environments

The simulation results presented in Chapter 5 suggest that the evolutionary process might be controlled by the trade-off between the fitness costs associated to the adaptation time and phenotypic robustness. A simple theoretical framework can be developed to further analyze the extent of these fitness costs defining the selected adaptation strategy, analyzing the explicit effect of the population size $N$, mutation rate $u$, mutation step-size $M$, environmental switching frequency $\nu$, and selection pressure $s_t$.

Given the evolutionary parameters, metrics for the expected adaptation time and robustness need to be developed. The adaptation time through epigenetic switching can be approached using the stochastic transition rate between bistable phenotypes (see Section 3.3.1) to estimate the expected number of epimutants each generation.
Then, how the frequency of “epimutant” individuals changes after an environmental transition by the selection and genetic drift forces needs to be analyzed. This metric will be directly affected by the population size $N$ and selection pressure $s_t$, as well as the specific bistable genotype. The adaptation time via genetic mutations can be estimated from a drift-diffusion model for a given fitness landscape (i.e. genotype-phenotype mapping). This metric will depend on all evolutionary parameters, except the environmental switching frequency $\nu$ (i.e. $N$, $u$, $M$, and $s_t$). And the phenotypic robustness of any genotype (either bistable or monostable) can be estimated considering both the effect of biochemical noise and de novo mutations. The biochemical noise effect can be estimated using the Finite State Projection (FSP) algorithm, which calculates a probability distribution dynamics given an initial condition and a state reaction matrix. This will depend on the selected genotype and the distribution of phenotypes passing from generation to generation, given the population size $N$ and selection pressure $s_t$. The effect of de novo mutations can be calculated by analyzing the mutational cloud associated to the selected genotype, which depends on the population size $N$, mutation rate $u$, and mutation step-size $M$. Then the expected fitness of this mutational cloud can be calculated considering the phenotypic landscape surrounding the selected genotype.

6.3 Other adaptation strategies for fluctuating environments

I verified that our observations were robust to many alternative model assumptions (see Section 3.4 and Figure 5.10). Nevertheless, our simple evolutionary model of a haploid, asexual population Schiffels et al. (2011) omits some features of the evolutionary process. For example, variable population size, diploid genetics, sexual reproduction and linkage disequilibrium could all affect the evolutionary dynamics and selection of epigenetic switches in a fluctuating environment.

Our evolutionary model also fixed the mutation rate ($u$) and step-size ($M$), which
imposes a mutational load when the population is adapted to a constant environment. Future simulations could allow natural selection to mutate and tune these parameters, which might favor genetic adaptation over epigenetic switching Desai and Fisher (2011). Similarly, our model did not consider the case where mutations (e.g. adaptive mutation) or biophysical parameters (e.g. phenotypic plasticity) directly respond to changes in the environment. Including a sensing mechanism, with the associated cost, is a natural extension for our model that needs to be explored.

Mutations were assumed continuously increase or decrease the biochemical parameters. This overlooked an important class of mutations, such as indels (i.e. rapid loss or abrupt change of function), gene duplication, and gene recruitment, which could abruptly change the topology of the gene network.

We considered the simplest genetic circuit that can exhibit epigenetic switching. However, alternative gene regulatory networks could generate different dynamics and phenotypes that are even better adapted to the fluctuating environment. For example, adding a negative feedback loop could reduce gene expression noise Becskei and Serrano (2000); Rosenfeld et al. (2002) or generate oscillations Novak and Tyson (2008); Tsai et al. (2008). An oscillatory gene circuit (e.g. circadian clock) might anticipate and respond to an environment that fluctuates regularly (e.g. day/night). Future research will explore more complicated gene regulatory circuits to understand the specific environmental dynamics and evolutionary conditions that favor oscillation versus epigenetic switching in the context of genetic adaptation. This should be of broad relevance to evolutionary biologists and systems biologists.

6.4 Considerations for the study of epigenetic switches in experimental systems

In experimental systems is usually difficult to determine values of the biophysical parameters with precision; then, biological bistable systems have been mostly detected
by the presence of a bimodal distribution of phenotypes in the population. Neverthe-
less, as observed in our computational experiments, in fast fluctuating environments
a population with a bistable genotype can appear unimodal even if its stationary
distribution is clearly bimodal given a relatively low rate of epimutation, a strong
selection pressure and the underlying population dynamics (see example in figure
3). In the experiments done by Balaban et al. (2004) in bacterial persistence and
by Beaumont et al. (2009) in bacterial colony morphology switching, the stochastic
epigeneric switching strategy was detected given the strong population bottlenecks
associated to the environmental transitions, but this might not always be the case.
Other approaches can help to detect this adaptation strategy in more general scenar-
ios. Lambert and Kussell (2015) have shown experimentally and theoretically that
selective pressures on a population, as those imposed by a fluctuating environment,
can be efficiently quantified by measurements on the surviving lineages; and more
recently, Cerulus et al. (2016) used life-history traits of cellular growth to show that
high single-cell variance in growth rate can be beneficial for the population, and that
this benefit depends on the epigenetic inheritance of the growth rate between mother
and daughter cells. Analogously, our analysis has shown that characterizing surviv-
ing lineages might be necessary to properly detect the selected adaptation strategy
in a population, particularly epigenetic switching. In general, this suggests that by
only measuring temporal snapshots of the population, the contribution of epigenetic
switching might be overlooked, and that future experimental studies need to consider
also the lineage information.
Appendix A

Stochastic dynamics of early transcriptional response – PIPELINE

All the calculations and analysis were performed in Matlab, and the code is presented accordingly.

A.1 Data

The experimental or synthetic data sample for each time measurement needs to be in a matrix form \( Y \) where each row \( (i = \{1, 2, 3\}) \) represents a promoter state \((p_{ON} + 1)\) and each column \((j = \{1, 2, ..., m_{MAX}\})\) the number of free mRNA molecules in the cell \((m + 1)\). A maximum mRNA number \((m_{MAX} = 300\) in all shown examples) needs to be defined (see Section 2.3.1). Data files must be named \( myData_[myGene][[N],[mMAX]].mat \), where \([myGene]\) is the gene name, \([N]\) is ‘2S’ for 2-states model, and \([mMAX]\) is the chosen \( m_{MAX} \) value (e.g. \( myData_Fos(2S,300).mat \)); here, the data is in a structure \( Y \) with \( Y.t00 \) having the data for the \( uninduced \) conditions, and \( Y.t## \) for the data for ## minutes after stimulus.
A.2 Define mathematical model

In order to accelerate the construction of the state reaction matrix for each set of biophysical parameters $A(\theta)$ during the Metropolis Random Walk (MRW) implementation, the instructions to fill the few elements different from zero are pre-defined given the 2-states model and $m_{MAX} = 300$ limit using the `STRUCT_TransM_2S` function:

```matlab
function [] = STRUCT_TransM_2S(mMAX)
    % Vector of states
    i = 1;
    for p = 0:2
        for m = 0:mMAX
            x(i).species = [p; m];
            i = i + 1;
        end
    end
    clear i p m

    % Transition matrix cases
    k = zeros(5,1); % Indexes
    for i = 1:length(x)
        for j = 1:length(x)
            nu = x(j).species - x(i).species;
            % For all i==j:
            if(nu==[0,0])
                k(1) = k(1)+1;
                myReaction.allNeg(k(1),:) = [i j];
                % If "promoter activation" occurs:
            elseif(nu==[1,0])
                k(2) = k(2)+1;
                myReaction.kON(k(2),:) = [i j];
                % If "promoter deactivation" occurs:
            elseif(nu==[-1,0])
                % Other cases
            end
        end
    end
end
```
\begin{verbatim}

k(3) = k(3)+1;
myReaction.kOFF(k(3),:) = [i j];  

\texttt{elseif\(\text{nu}==\{0\,1\}\)}
\texttt{k(4) = k(4)+1;}
\texttt{myReaction.mus(k(4),:) = [i j];  }

\texttt{elseif\(\text{nu}==\{0\,-1\}\)}
\texttt{k(5) = k(5)+1;}
\texttt{myReaction.d(k(5),:) = [i j];  }
\end{verbatim}

\texttt{end  
end  
end  
}

\texttt{clear \, i \, j \, k \, \text{nu}  

\% Functions  

\% % LEAVING \([xJ]\) STATE \% %
myPropensity.allNeg = @(Par,xJ) [...  
\quad \texttt{(Par.kON\*(2-xJ\text{.species\,(1))})...}
\quad \texttt{(Par.kOFF\*xJ\text{.species\,(1))}...}
\quad \texttt{(Par.mu0\*(2-xJ\text{.species\,(1))}...}
\quad \texttt{(Par.mu\*xJ\text{.species\,(1))}...}
\quad \texttt{(Par.d\*xJ\text{.species\,(2))});  

\% % ENTERING \([xJ]\) STATE \% %
\%
\%

\% Promoter activation \((pOFF \rightarrow pON)\):
myPropensity.kON = @(Par,xJ) [Par.kON\*(2-xJ\text{.species\,(1))}];

\% Promoter deactivation \((pON \rightarrow pOFF)\):
myPropensity.kOFF = @(Par,xJ) [Par.kOFF\*xJ\text{.species\,(1))};

\% mRNA synthesis \((m \rightarrow m+1)\):
myPropensity.mus = @(Par,xJ) [...]  
\quad \texttt{[(Par.mu0\*(2-xJ\text{.species\,(1))})...}
\quad \texttt{+(Par.mu\*xJ\text{.species\,(1))});  

\% mRNA degradation \((m \rightarrow m-1)\)
myPropensity.d = @(Par,xJ) [Par.d\*xJ\text{.species\,(2))];

\%

\%

\% Exclude cases when mRNA molecule number is in the limit
\% of the matrix; if mRNA is already the maximum, no
\% synthesis can occur:
myPropensity.Exc.allNeg = @(Par,xJ) [...  
\quad \texttt{(Par.mu0\*(2-xJ\text{.species\,(1))})...}
\quad \texttt{(Par.mu\*xJ\text{.species\,(1))});

\% Save

\end{verbatim}
And then, each time the kinetic parameters change, construct the state reaction matrix using the **DATA_TransM** function:

```matlab
% DATA_TransM : Calculate the model’s state reaction matrix
% given the chosen kinetic parameters.

% A = DATA_TransM(mMAX, Par)
% mMAX : Maximum mRNA number to consider.
% Par : Structure with the model parameters.
% .kON : Promoter activation rate (OFF->ON)
% .kOFF : Promoter deactivation rate (ON->OFF)
% .mu0 : mRNA synthesis rate of promoter in OFF state
% .mu : mRNA synthesis rate of promoter in ON state
% .d : mRNA degradation rate

function A = DATA_TransM(mMAX, Par)
    load( cat(2, 'myTransM(2S, ', num2str(mMAX), ').mat')
    % Transition matrix
    A = zeros(length(x), length(x));
    for myI = 1:length(myNames)
        myR = myNames{myI};
        for myJ = 1:length(myReaction.(myR))
            i = myReaction.(myR)(myJ,1);
            j = myReaction.(myR)(myJ,2);
            A(j,i) = myPropensity.(myR)(Par,x(i));
        end
    end
end
```

**A.3 Calculate probability distributions**

As explained in Section 2.3.1, given the state reaction matrix $A$, the stationary distribution ($P_U$) is calculated for the *uninduced* biophysical parameters ($\theta_U$) using the **DATA_Pss** function, and the probability distribution dynamics with parameters $\theta_S$ assuming $P_U$ as the initial condition using the **DATA_Pxt** function:
% DATA_Pss : Calculate the stationary distribution vector
% given a transition matrix.
%  
%  Pss = DATA_Pss(A)
%  A : State reaction matrix describing the model.
%  Pss : Stationary distribution with states as
%        described in A structure.

function Pss = DATA_Pss(A)

    % Find the eigenvector with eigenvalue 0:
    [V, lambda] = eigs(A, 1, 0);
    % Normalize eigenvector:
    Pss = abs(V/sum(V));
end

% DATA_Pxt : Calculate the probability distribution at time
% t given a transition matrix and some initial
% condition.
%  
%  Pt = DATA_Pxt(A,P0,t)
%  A : State reaction matrix describing the model.
%  P0 : Initial probability distribution with states as
%        described in A structure.
%  t : Time (e.g. 5).
%  Pt : Probability distribution at time t with states
%        as described in A structure.

function Pt = DATA_Pxt(A,P0,t)

    Pt = expm(A*t)*P0;
end

A.4 Compare model & data

In order to evaluate the model (i.e. particular set of parameters \( \{ \theta_U, \theta_S \} \)), the log-
likelihood of observing the data \( Y \) given the probability distributions associated to
the model as explained in Section 2.3.2 using the DATA_logL function:

% DATA_logL : Calculate the log-likelihood of observing the
% data Y given the probability distribution P.
%  
%  L = DATA_logL_2S(P,Y)
% $P$: Probability matrix where $P(i,j)$ is the probability of having the $(i-1)$ active promoters and exactly $(j-1)$ mRNA molecules.
% $Y$: Observed data matrix where $Y(i,j)$ is the number of individuals in the population with $(i-1)$ active promoters and exactly $(j-1)$ mRNA molecules.
% *NOTE: $P$ and $Y$ must have the same dimensions.

% $L$: log-likelihood of $Y$ given $P$, normalized by number of observations, excluding the multinomial coefficient.

define function $L = \text{DATA}\_\log L(P,Y)$

% Imposing a minimal value for the probability values to avoid numerical errors:
$P(|P<1e-100|) = 1e-100$;
$L = Y \times \log(P)$;
$L = \text{sum}(\text{sum}(L))/\text{sum}(\text{sum}(Y))$;

To translate the probability distributions from vector form $Pv$, as the output from $\text{DATA}\_\text{Pss}$ and $\text{DATA}\_\text{Pxt}$ functions, to a matrix form $Pm$, as the data structure, simply apply:

$Pm = \text{vec2mat}(Pv, \text{length}(Pv)/3)$;

A.5 Metropolis Random Walk algorithm

To run the Metropolis Random Walk algorithm over a particular model and experimental data, use the $\text{SIM}\_\text{mrw}$ function:

% $SIM\_\text{mrw}$: Run Metropolis Random Walk algorithm given a model and observed data $Y$.
% $[] = SIM\_\text{mrw}(\text{myGeneModel}, \text{Par}, \text{mrw}, \text{cont})$
% myGeneModel: Gene model to simulate (e.g. '$[\text{myGene}([N],[\text{mMAX}])$', where '$[N]$' corresponds to the 2-state model, e.g. '2S', and '$[\text{mMAX}]$' is the limit for the mRNA molecule number, e.g. 300).
% Par: Structure with the kinetic parameters.
% .kON: Promoter activation rate (OFF->ON)
\% \texttt{kOFF} : Promoter deactivation rate (ON->OFF)
\% \texttt{mu0} : mRNA synthesis rate of promoter in OFF state
\% \texttt{mu} : mRNA synthesis rate of promoter in ON state
\% \texttt{d} : mRNA degradation rate
\% \texttt{mrw} : Algorithm parameters.
\% \texttt{I} : Number of iterations (e.g. 100000)
\% \texttt{s} : Random number generator seed (e.g. 1)
\% \texttt{([U/S]).(par)} : [sigma, min, max] (e.g. [1e-5,1e-3,10])
\% \texttt{[U/S]} : \textit{S} if the parameter changes between 'basal'
\% \texttt{and 'stimulus'} ; \textit{U}, otherwise.
\% \texttt{sigma} : Covariance of (\texttt{par}) to calculate parameter
\% perturbations.
\% \texttt{min, max} : Initial value range of (\texttt{Par}).
\% \texttt{P} : OUTPUT – Parameters per iteration
\% \texttt{L} : OUTPUT – Log-likelihood per iteration
\% \texttt{e} : OUTPUT – Maximum error observed in the
\% distribution estimation
\% \texttt{cont} : If 1, continue stopped simulation saved on file
\% 'TEMP_MRW_{myGene}([N],[maxM],
\% [fieldnames(ParS)](s{mrw.s}).mat'
\% OUTPUT FILE : 'MRW_{myGene}([N],[mMAX],
\% [fieldnames(ParS)](s{mrw.s}).mat'

\begin{verbatim}
function [] = SIM_mrw(myGeneModel, Par, mrw, cont)
    \% Model (N), maximum mRNA number to consider (mMAX),
    \% and data (Y):
    load(cat(2, 'myData_', myGeneModel, '.mat'), 'N', 'mMAX', 'Y');

    % Define indexes:
    myI = [];
    I1 = fieldnames(mrw.U);
    for i = 1:length(I1)
        myI = [myI; {'U', I1{i}, 1}];
    end
    I1 = fieldnames(mrw.S);
    psN = '.'; \% ParS fieldnames to name results.
    for i = 1:length(I1)
        myI = [myI; {'S', I1{i}, 1}; {'S', I1{i}, 2}];
        psN = cat(2, psN, I1{i});
    end
    clear i I1

    % Specify time points (myT) and sample size (myS):

\end{verbatim}
```matlab
myT = fieldnames(Y);
myS = zeros(1, length(myT));
for t = 1:length(myT)
    myS(t) = sum(sum(Y.(myT{t})));
end
clear t
if (cont==1)
    load([cat(2, 'TEMP.MRW.', myGeneModel, '(', psN, ')', '(', s, ')', num2str(mrw.s), ').mat']);
else
    % (1) Define random number generator:
    rng(mrw.s, 'twister');
    r.p0 = rand(size(myI,1),1);
    for i = 1:size(myI,1)
        Z(i) = mrw.(myI{i,1}).(myI{i,2})(myI{i,3},1);
    end
    r.Pe = mvnrnd(zeros(mrw.I, length(Z)), Z);
    r.tL = rand(mrw.I,1);
    clear i Z
    % (2) Initialize system:
    mrw.P = zeros(mrw.I, size(myI,1));
    mrw.L = zeros(mrw.I, length(myT));
    mrw.e = zeros(mrw.I,1);
    % Initial parameter set:
    for i = 1:size(myI,1)
        if(strcmp(myI{i,1}, 'U'))
            Par.(myI{i,2}) = mrw.U.(myI{i,2})(2) + ... 
            (r.p0(i)* (mrw.U.(myI{i,2})(3)... 
            -mrw.U.(myI{i,2})(2)));
            mrw.P(1,i) = Par.(myI{i,2});
        else
            ParS.(myI{i,2})(myI{i,3}) = ... 
            mrw.S.(myI{i,2})(myI{i,3},2)... 
            + (r.p0(i) ... 
            *(mrw.S.(myI{i,2})(myI{i,3},3)... 
            -mrw.S.(myI{i,2})(myI{i,3},2)));
            mrw.P(1,i) = ParS.(myI{i,2})(myI{i,3});
        end
    end
    [m, l, e] = SIM_logL(Y,N,mMAX,Par,ParS);
end
for t = 1:length(myT)
```

mrw.L(1,t) = myS(t)*l.(myT{t});

mrw.e(1) = max(mrw.e(1), abs(e.(myT{t})));
end
clear itemle
j0 = 2;
end

% (3) Iterate:
for j = j0:mrw.I
    mrw.P(j,:) = mrw.P(j-1,:);
    mrw.L(j,:) = mrw.L(j-1,:);
    mrw.e(j) = mrw.e(j-1,:);
    % Alternative parameter set:
    myP = mrw.P(j,:) + r.Pe(j,:);
    if (min(myP)>=1e-8)
        for i = 1:size(myI,1)
            if (strcmp(myI{i,1},'U'))
                Par.(myI{i,2}) = myP(i);
            else
                ParS.(myI{i,2})(myI{i,3}) = myP(i);
            end
        end
    end
    % Generate proposal:
    [m,l,e] = SIM_logL(Y,N,mMAX,Par,ParS);
    myL = zeros(1,length(myT));
    myE = 0;
    for i = 1:length(myT)
        mL(i) = myS(i)*l.(myT{i});
        myE = max(myE, abs(e.(myT{i})));
    end
    clear itemle
    % If proposal is accepted, update system:
    if (r.tL(j) <= 
        min(1, exp(sum(myL)-sum(mrw.L(j,:)))))
        mrw.P(j,:) = myP;
        mrw.L(j,:) = myL;
        mrw.e(j) = myE;
    end
end
% Save progress:
if (mod(j,100)==0)
    j0 = j + 1
    save(cat(2,'TEMP_MRW',myGeneModel,...
clear j myP myL myE

% (4) Save:
save(cat(2,'MRW_','myGeneModel','(',psN,')','(s',num2str(mrw.s),')',mat'),...
'mrw','j0','r','Par','ParS');
delete(cat(2,'TEMPMRW_','myGeneModel','(',psN,')','(s',num2str(mrw.s),')',mat'));

% SIM_logL : Calculate the log-likelihood of observing
% the data Y for a given set of kinetic
% parameters.
% [P,L,eps] = SIM_logL(Y,N,mMAX,Par,ParS)
% Y : Observed data matrix where Y(i,j) is the number of
% individuals in de population with (i−1) active
% promoters and exactly (j−1) mRNA molecules.
% N : Model ('2S').
% mMAX : Maximum mRNA number to consider.
% Par : Structure with the kinetic parameters.
% .kON : Promoter activation rate (OFF→ON)
% .kOFF : Promoter deactivation rate (ON→OFF)
% .mu0 : mRNA synthesis rate of promoter in OFF state
% .mu : mRNA synthesis rate of promoter in ON state
% .d : mRNA degradation rate
% ParS : Structure with the kinetic parameters sensitive
% to stimulus as elements, and a 2-vector with the
% value of the given parameter in basal and stimulus
% conditions, [Basal,Stimulus] (e.g. ParS.kON =
% [0.0051,0.1373]).
% P : Probability matrix matrix where P(i,j) is the
% probability of having (i−1) active promoters and
% exactly (j−1) mRNA molecules.
% L : log-likelihood of Y given P, normalized by number
% of observations.
% eps : FSP algorithm estimation error.

function [P,L,eps] = SIM_logL(Y,N,mMAX,Par,ParS)
% Parameters sensitive to stimulus:
pS = fieldnames(ParS);
% Time points:
myT = fieldnames(Y);
% Basal conditions:
for p = 1:length(pS)
    Par.(pS{p}) = ParS.(pS{p})(1);
end
A = DATA_TransM(N,mMAX,Par);
% Probability distribution vector:
Pt = DATA_Pss(A);
% Probability distribution matrix:
P.t00 = DATA_P2M(N,Pt);
% Calculate the log–likelihood:
L.t00 = DATA_logL(P.t00,Y.t00);
% Calculate estimation error:
eps.t00 = 1–sum(abs(Pt));
% Stimulus conditions:
for p = 1:length(pS)
    Par.(pS{p}) = ParS.(pS{p})(2);
end
A = DATA_TransM(N,mMAX,Par);
At5 = expm(A*5);
T = zeros(1,length(myT));
for t = 2:length(myT)
    tt = regexp(myT{t},'\.*d*','match');
    T(t) = str2double(tt{1});
    temp = T(t) – T(t-1);
    while(temp > 0)
        % Probability distribution vector:
        Pt = At5*Pt;
        temp = temp - 5;
    end
    % Probability distribution matrix:
P.(myT{t}) = vec2mat(Pt,length(Pt)/3);
% Calculate the log–likelihood:
L.(myT{t}) = DATA_logL(P.(myT{t}),Y.(myT{t}));
% Calculate FSP algorithm estimation error:
eps.(myT{t}) = 1–sum(Pt);
if(eps.(myT{t}) > 0.001)
cat(2,'CAUTION: Large FSP algorithm estimation error',...'
estimation_error(',...num2str(eps.(myT{t})),',')')
"CAUTION: Large FSP algorithm estimation error",...
"estimation_error","num2str(eps.(myT{t})),"/Private/Temp/166"
In all examples in Section 2.4, the algorithm parameters are:

```matlab
% Data & model frame:
myGeneModel = 'Fos(2S,300)';
% Fix kinetic parameters:
Par.d = 0.0462;
% Metropolis Random Walk algorithm parameters:
mrw.I = 35000;
mrw.s = 1;    % Changes with each replica.
```

And for each model, the specific parameters sensitive to stimulus are specified in the `mrw.S` structure, and the rest in `mrw.U` with the covariance sigma, and the range for the random initial value:

```matlab
% If parameter is sensitive to stimulus:
mrw.S.kON = [1e-5, 1e-6, 0.01;...  
             1e-5, 1e-4, 1];
mrw.S.kOFF = [1e-5, 1e-4, 1;...  
              1e-5, 1e-6, 0.01];
mrw.S.mu   = [1e-3, 1e-4, 1;...  
             1e-3, 1e-3, 10];
% If parameter is insensitive to stimulus:
mrw.U.kON  = [1e-5, 1e-5, 0.1;...  
mrw.U.kOFF = [1e-5, 1e-5, 0.1];
mrw.U.mu0  = [1e-5, 1e-5, 0.1];
mrw.U.mu   = [1e-3, 1e-3, 10];
```

### A.6 Synthetic data

Given the probability distributions describing a model with some biophysical parameters, we can generate a random sample per time point using the `DATA_Xrs` function:

```matlab
% DATA_Xrs : Generate a random sample of size n from
% given a probability distribution matrix P.
```
% Xrs = DATA_Xrs(P,n,s)
% P : Probability distribution matrix where P(i,j) is
% the probability of having (i-1) active promoters
% and exactly (j-1) mRNA molecules.
% n : Sample size, i.e. number of observations to generate.
% s : Random number generator seed (e.g. 11)
% Xr : Random sample from the given probability distribution P.
% * NOTE: Xr has the same dimensions that P.

function Xrs = DATA_Xrs(P,n,s)
% Probability distribution vector:
Pv = vec2mat(P,1);
% Cumulative distribution:
Pc = Pv;
for i = 2:length(Pv)
Pc(i) = Pc(i) + Pc(i-1);
end
% Correct from small numeric errors:
Pc(length(Pc)) = 1;
% Define random number generator:
rng(s,'twister');
% Sample:
Xrs = zeros(length(Pv),1);
for i = 1:n
    r = sum([Pc<rand()])+1;
    Xrs(r) = Xrs(r) + 1;
end
Xrs = vec2mat(Xrs,size(P,2));

Then, once defined the time point measurements of interest (T) and the sample size for each of these (n), the synthetic data for a model with probability distributions \{P_U,P_{\tau}\} is:

% Time points:
srs.t = {'t00','t05','t15'};
% 't00' corresponds to the uninduced condition.
% Sample size:
srs.n = [200 200 200];
% Random number generator seed:
srs.s = 1;
% Generate random sample per time point:
for t = 1:length(srs.t)
    x.(srs.t{t}) = ...
    DATA_Xrs(P.(srs.t{t}), srs.n(t), (srs.s*10)+t);
end
% NOTE: P. (t) is the probability distribution
% per time point t.
Appendix B

Previous studies on stochastic dynamics of 2-states models of gene expression

<table>
<thead>
<tr>
<th>Zenklusen et al. (2008)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Experimental system:</strong> smFISH (yeast)</td>
</tr>
<tr>
<td><strong>Model:</strong> Constitutive &amp; 2-states</td>
</tr>
<tr>
<td><strong>Approach:</strong> Monte Carlo simulation &amp; analytical distributions</td>
</tr>
<tr>
<td><strong>Parameter fitting:</strong> $\chi^2$ test</td>
</tr>
<tr>
<td><strong>Synthesis delay:</strong> Yes</td>
</tr>
<tr>
<td><strong>Comments:</strong> The authors explored bursting in yeast by analyzing the mRNA expression distribution and its associated kinetic parameters of a set of canonical genes. Both constitutive and bursting behaviors were observed.</td>
</tr>
<tr>
<td>So et al. (2011)</td>
</tr>
<tr>
<td>---</td>
</tr>
<tr>
<td><strong>Experimental system:</strong></td>
</tr>
<tr>
<td><strong>Model:</strong></td>
</tr>
<tr>
<td><strong>Approach:</strong></td>
</tr>
<tr>
<td><strong>Parameter fitting:</strong></td>
</tr>
<tr>
<td><strong>Synthesis delay:</strong></td>
</tr>
<tr>
<td><strong>Comments:</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Senecal et al. (2014)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Experimental system:</strong></td>
<td>smFISH + immunofluorescence (human tissue)</td>
</tr>
<tr>
<td><strong>Model:</strong></td>
<td>2-states &amp; 3-states</td>
</tr>
<tr>
<td><strong>Approach:</strong></td>
<td>Estimate the probability that a promoter is active assuming a binomial distribution of TSs; use FSP to describe nascent mRNA distribution &amp; find ”good” fitted parameters; choose the parameters whose mature mRNA simulated distribution (Gillespie) better fit experimental data.</td>
</tr>
<tr>
<td><strong>Parameter fitting:</strong></td>
<td>Sum of square residuals &amp; the Kullback-Leibler divergence.</td>
</tr>
<tr>
<td><strong>Synthesis delay:</strong></td>
<td>Yes</td>
</tr>
<tr>
<td><strong>Comments:</strong></td>
<td>The authors explored the effect of transcription factors regulation over the early response gene c-Fos; they concluded that TF concentration modulates the burst frequency of c-Fos.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Bahar Halpern et al. (2015)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Experimental system:</strong></td>
<td>smFISH (mouse liver)</td>
</tr>
<tr>
<td><strong>Model:</strong></td>
<td>Constitutive &amp; 2-states</td>
</tr>
<tr>
<td><strong>Approach:</strong></td>
<td>Fit $k_{ON}$ to mRNA distributions using measured burst parameters</td>
</tr>
<tr>
<td><strong>Parameter fitting:</strong></td>
<td>Mean squared errors</td>
</tr>
<tr>
<td><strong>Synthesis delay:</strong></td>
<td>No</td>
</tr>
<tr>
<td><strong>Comments:</strong></td>
<td>The authors explored bursting in mouse liver by analyzing the mRNA expression distribution correcting by the natural liver polyploidy; the authors showed that coordination of bursts parameters and polyploidy can help to regulate gene expression noise.</td>
</tr>
</tbody>
</table>
### Bintu et al. (2016)

<table>
<thead>
<tr>
<th><strong>Experimental system:</strong></th>
<th>Time-lapse microscopy &amp; H2B-citrine</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Model:</strong></td>
<td>2-states &amp; 3-states</td>
</tr>
<tr>
<td><strong>Approach:</strong></td>
<td>ODE analytical solution</td>
</tr>
<tr>
<td><strong>Parameter fitting:</strong></td>
<td>Nonlinear least square fit (MATLAB)</td>
</tr>
<tr>
<td><strong>Synthesis delay:</strong></td>
<td>No</td>
</tr>
<tr>
<td><strong>Comments:</strong></td>
<td>The authors followed and characterized the regulatory dynamics of four chromatin silencers; they observed in all cases that silencing and reactivation occurred in all-or-none events, and different regulators moderate only the probability of these events.</td>
</tr>
</tbody>
</table>

### Sepulveda et al. (2016)

<table>
<thead>
<tr>
<th><strong>Experimental system:</strong></th>
<th>smFISH + immunofluorescence (lysogeny maintenance promoter of bacteriophage lambda)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Model:</strong></td>
<td>4-promoter states x 2-activity states</td>
</tr>
<tr>
<td><strong>Approach:</strong></td>
<td>Quasi-equilibrium assumption, FSP to estimate the stationary distribution</td>
</tr>
<tr>
<td><strong>Parameter fitting:</strong></td>
<td>Maximum likelihood</td>
</tr>
<tr>
<td><strong>Synthesis delay:</strong></td>
<td>No</td>
</tr>
<tr>
<td><strong>Comments:</strong></td>
<td>The authors explored the role and dynamics of promoter configurations on mRNA expression and cell-to-cell variability. They concluded that the switching between promoter states is faster than the mRNA lifetime and that the alleles of the gene are independently active.</td>
</tr>
</tbody>
</table>

### Our model

<table>
<thead>
<tr>
<th><strong>Experimental system:</strong></th>
<th>smFISH (mouse neurons)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Model:</strong></td>
<td>2-states</td>
</tr>
<tr>
<td><strong>Approach:</strong></td>
<td>CME, stationary distributon &amp; FSP to characterize the distribution of mRNA expression in the model and dynamics after stimulus perturbation.</td>
</tr>
<tr>
<td><strong>Parameter fitting:</strong></td>
<td>Bayesian approach (Metropolis algorithm)</td>
</tr>
<tr>
<td><strong>Synthesis delay:</strong></td>
<td>No</td>
</tr>
<tr>
<td><strong>Comments:</strong></td>
<td>We take into account the intrinsic stochasticity, as well as the dynamic response, of the Fos mRNA expression in single-neurons to characterize the kinetic parameters associated to the regulatory system. The implemented Bayesian approach allows us to overcome the limitation of a small, noisy sample, and evaluate the parameter identifiability and reproducibility of our model.</td>
</tr>
</tbody>
</table>
Appendix C

Previous studies on evolution of epigenetic switches

<table>
<thead>
<tr>
<th>Jablonka et al. (1995)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Evolutionary strategies:</strong></td>
</tr>
<tr>
<td><strong>Comparison method:</strong></td>
</tr>
<tr>
<td><strong>Analytical/Numerical:</strong></td>
</tr>
<tr>
<td><strong>Phenotypes:</strong></td>
</tr>
<tr>
<td><strong>Environment:</strong></td>
</tr>
<tr>
<td><strong>Fitness:</strong></td>
</tr>
<tr>
<td><strong>Selection:</strong></td>
</tr>
<tr>
<td>** Explicit cost:**</td>
</tr>
<tr>
<td><strong>Population size:</strong></td>
</tr>
<tr>
<td><strong>Generations:</strong></td>
</tr>
<tr>
<td><strong>Comments:</strong></td>
</tr>
</tbody>
</table>
### Lachmann and Jablonka (1996)

<table>
<thead>
<tr>
<th>Evolutionary strategies:</th>
<th>Inducible switching; stochastic switching</th>
</tr>
</thead>
<tbody>
<tr>
<td>Comparison method:</td>
<td>Average growth rate</td>
</tr>
<tr>
<td>Analytical/Numerical:</td>
<td>Analytical</td>
</tr>
<tr>
<td>Phenotypes:</td>
<td>2-discrete</td>
</tr>
<tr>
<td>Environment:</td>
<td>Periodic (2 states)</td>
</tr>
<tr>
<td>Fitness:</td>
<td>Symmetric</td>
</tr>
<tr>
<td>Selection:</td>
<td>Differential growth rate</td>
</tr>
<tr>
<td>Explicit cost:</td>
<td>None</td>
</tr>
<tr>
<td>Population size:</td>
<td>Growing (discrete)</td>
</tr>
<tr>
<td>Generations:</td>
<td>Non-overlapping</td>
</tr>
<tr>
<td>Comments:</td>
<td>The authors explored the optimal values for the transition rates under fluctuating environments; they concluded that for non-inducible systems, the optimal rate for random transitions is around the frequency of the environmental fluctuations.</td>
</tr>
</tbody>
</table>

### Thattai and Oudenaarden (2004)

<table>
<thead>
<tr>
<th>Evolutionary strategies:</th>
<th>Inducible switching; stochastic switching</th>
</tr>
</thead>
<tbody>
<tr>
<td>Comparison method:</td>
<td>Average growth rate</td>
</tr>
<tr>
<td>Analytical/Numerical:</td>
<td>Analytical</td>
</tr>
<tr>
<td>Phenotypes:</td>
<td>2-discrete</td>
</tr>
<tr>
<td>Environment:</td>
<td>Periodic &amp; random (2 states)</td>
</tr>
<tr>
<td>Fitness:</td>
<td>Symmetric</td>
</tr>
<tr>
<td>Selection:</td>
<td>Differential growth rate</td>
</tr>
<tr>
<td>Explicit cost:</td>
<td>None</td>
</tr>
<tr>
<td>Population size:</td>
<td>Growing (continuous)</td>
</tr>
<tr>
<td>Generations:</td>
<td>Continuous time (ODEs)</td>
</tr>
<tr>
<td>Comments:</td>
<td>The authors considered that the transitions between phenotypic states depend on the environment and explored under which circumstances a transition rate to the unfit state different to zero will be selected; they concluded that if the transition to the fit state is fast enough short induction delay, a homogeneous population will be always favored.</td>
</tr>
</tbody>
</table>
### Kussell and Leibler (2005)

| **Evolutionary strategies:** | Inducible switching; stochastic switching |
| **Comparison method:**       | Average growth rate                         |
| **Analytical/Numerical:**    | Analytical                                   |
| **Phenotypes:**              | n-discrete                                   |
| **Environment:**             | Random (n states)                            |
| **Fitness:**                 | No explicit assumption                       |
| **Selection:**               | Differential growth rate                     |
| **Explicit cost:**           | Sensing, diversity & induction delay costs   |
| **Population size:**         | Growing (continuous)                         |
| **Generations:**             | Continuous time (ODEs)                       |
| **Comments:**                | The authors compared inducible to stochastic transitions but taking into account the cost of sensing, the induction delay and the diversity cost imposed by the stochastic switching. The authors concluded that a sensor is only worth if the environment is highly uncertain, and the stochastic switching will be favored when the environment changes infrequently. |

### Kussell et al. (2005)

<p>| <strong>Evolutionary strategies:</strong> | Stochastic switching |
| <strong>Comparison method:</strong>       | Average growth rate |
| <strong>Analytical/Numerical:</strong>    | Both               |
| <strong>Phenotypes:</strong>              | 2-discrete         |
| <strong>Environment:</strong>             | Periodic; asymmetric (2 states) |
| <strong>Fitness:</strong>                 | Asymmetric         |
| <strong>Selection:</strong>               | Differential growth rate |
| <strong>Explicit cost:</strong>           | None               |
| <strong>Population size:</strong>         | Growing (continuous) |
| <strong>Generations:</strong>             | Continuous time (ODEs) |
| <strong>Comments:</strong>                | The authors considered only stochastic transitions, and they observed that the type of environmental changes determines the strategy to be used. |</p>
<table>
<thead>
<tr>
<th><strong>Wolf et al. (2005)</strong></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Evolutionary strategies:</strong></td>
<td>Fixed; inducible switching; stochastic switching</td>
</tr>
<tr>
<td><strong>Comparison method:</strong></td>
<td>Average growth rate</td>
</tr>
<tr>
<td><strong>Analytical/Numerical:</strong></td>
<td>Both</td>
</tr>
<tr>
<td><strong>Phenotypes:</strong></td>
<td>$n$-discrete (focus on $n = 2$)</td>
</tr>
<tr>
<td><strong>Environment:</strong></td>
<td>Random; asymmetric ($n$ states)</td>
</tr>
<tr>
<td><strong>Fitness:</strong></td>
<td>Asymmetric</td>
</tr>
<tr>
<td><strong>Selection:</strong></td>
<td>Differential growth rate</td>
</tr>
<tr>
<td><strong>Explicit cost:</strong></td>
<td>None</td>
</tr>
<tr>
<td><strong>Population size:</strong></td>
<td>Growing (discrete)</td>
</tr>
<tr>
<td><strong>Generations:</strong></td>
<td>Non-overlapping</td>
</tr>
<tr>
<td><strong>Comments:</strong></td>
<td>The authors considered more flexible adaptation strategies, going from ignoring the environment, a deterministic inducible response, stochastic inducible response, to pure stochastic switching. If no sensor exists, stochastic switching is always selected under the time-varying environmental conditions selected here, as well as if the detection of the sensor is bad or long induction delays exist.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Ribeiro (2008)</strong></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Evolutionary strategies:</strong></td>
<td>Inducible; stochastic switching (bistable genetic circuit)</td>
</tr>
<tr>
<td><strong>Comparison method:</strong></td>
<td>Invasion</td>
</tr>
<tr>
<td><strong>Analytical/Numerical:</strong></td>
<td>Simulations</td>
</tr>
<tr>
<td><strong>Phenotypes:</strong></td>
<td>Continuous (mechanistic)</td>
</tr>
<tr>
<td><strong>Environment:</strong></td>
<td>Random (2 states)</td>
</tr>
<tr>
<td><strong>Fitness:</strong></td>
<td>Symmetric</td>
</tr>
<tr>
<td><strong>Selection:</strong></td>
<td>Truncation selection</td>
</tr>
<tr>
<td><strong>Explicit cost:</strong></td>
<td>None</td>
</tr>
<tr>
<td><strong>Population size:</strong></td>
<td>Fixed (discrete; 1000 individuals)</td>
</tr>
<tr>
<td><strong>Generations:</strong></td>
<td>Non-overlapping</td>
</tr>
<tr>
<td><strong>Comments:</strong></td>
<td>The author modeled individual cells as toggle switches and explored the population behavior under a fluctuating environment, considering both inducible systems and pure stochastic switching. The author concluded that the optimal noise level depends on the environmental fluctuations, and as noise increases, the fitness increases too in fast fluctuating environments.</td>
</tr>
<tr>
<td><strong>Salathé et al. (2009)</strong></td>
<td></td>
</tr>
<tr>
<td>--------------------------</td>
<td>--------------------------</td>
</tr>
<tr>
<td><strong>Evolutionary strategies:</strong></td>
<td>Fixed; stochastic switching</td>
</tr>
<tr>
<td><strong>Comparison method:</strong></td>
<td>Invasion</td>
</tr>
<tr>
<td><strong>Analytical/Numerical:</strong></td>
<td>Simulations</td>
</tr>
<tr>
<td><strong>Phenotypes:</strong></td>
<td>2-discrete (x 2-modifier states)</td>
</tr>
<tr>
<td><strong>Environment:</strong></td>
<td>Periodic &amp; random (2 states)</td>
</tr>
<tr>
<td><strong>Fitness:</strong></td>
<td>Asymmetric</td>
</tr>
<tr>
<td><strong>Selection:</strong></td>
<td>Proportional selection scheme</td>
</tr>
<tr>
<td><strong>Explicit cost:</strong></td>
<td>None</td>
</tr>
<tr>
<td><strong>Population size:</strong></td>
<td>Infinite (sub-population frequencies)</td>
</tr>
<tr>
<td><strong>Generations:</strong></td>
<td>Non-overlapping</td>
</tr>
<tr>
<td><strong>Comments:</strong></td>
<td>Assuming an infinite population and following sub-population frequencies through generations, the authors explored the impact of asymmetric fitness landscapes. The authors concluded that with the fitness asymmetry over a certain threshold, unless the selection pressure is very strong in both environments, ignoring the environment becomes optimal over stochastic switching (with an optimal rate approximately equal to the environmental fluctuation frequency).</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Gaál et al. (2010)</strong></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Evolutionary strategies:</strong></td>
<td>Fixed; stochastic switching</td>
</tr>
<tr>
<td><strong>Comparison method:</strong></td>
<td>Average growth rate</td>
</tr>
<tr>
<td><strong>Analytical/Numerical:</strong></td>
<td>Analytical</td>
</tr>
<tr>
<td><strong>Phenotypes:</strong></td>
<td>2-discrete</td>
</tr>
<tr>
<td><strong>Environment:</strong></td>
<td>Periodic; asymmetric (2 states)</td>
</tr>
<tr>
<td><strong>Fitness:</strong></td>
<td>Asymmetric</td>
</tr>
<tr>
<td><strong>Selection:</strong></td>
<td>Differential growth rate</td>
</tr>
<tr>
<td><strong>Explicit cost:</strong></td>
<td>None</td>
</tr>
<tr>
<td><strong>Population size:</strong></td>
<td>Infinite (sub-population frequencies)</td>
</tr>
<tr>
<td><strong>Generations:</strong></td>
<td>Continuous time (ODEs)</td>
</tr>
<tr>
<td><strong>Comments:</strong></td>
<td>The authors observed that as the asymmetry in the environments increases, the selected strategy goes from the optimal stochastic switching population (where the transition rate is assumed equal in both directions) to an equally optimal non-switching and switching populations, to finally being optimal to ignore the environment, even if a local maximum still exists for a switching rate distinct to zero.</td>
</tr>
<tr>
<td>Visco et al. (2010)</td>
<td></td>
</tr>
<tr>
<td>---------------------</td>
<td></td>
</tr>
<tr>
<td><strong>Evolutionary strategies:</strong> Fixed; stochastic switching</td>
<td></td>
</tr>
<tr>
<td><strong>Comparison method:</strong> Average growth rate</td>
<td></td>
</tr>
<tr>
<td><strong>Analytical/Numerical:</strong> Analytical</td>
<td></td>
</tr>
<tr>
<td><strong>Phenotypes:</strong> 2-discrete</td>
<td></td>
</tr>
<tr>
<td><strong>Environment:</strong> Responsive (i.e. catastrophe rate depends on the population); random (one normal state &amp; instantaneous catastrophe)</td>
<td></td>
</tr>
<tr>
<td><strong>Fitness:</strong> Asymmetric</td>
<td></td>
</tr>
<tr>
<td><strong>Selection:</strong> Differential growth rate</td>
<td></td>
</tr>
<tr>
<td><strong>Explicit cost:</strong> None</td>
<td></td>
</tr>
<tr>
<td><strong>Population size:</strong> Growing (continuous)</td>
<td></td>
</tr>
<tr>
<td><strong>Generations:</strong> Continuous time (ODEs)</td>
<td></td>
</tr>
<tr>
<td><strong>Comments:</strong> The authors explored the selection of stochastic switching under a single environment with occasional and instantaneous catastrophic events whose rate depends on the population structure. The authors observed that stochastic switching strategy is favored by strong catastrophes, while non-switching by weak catastrophes.</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Liberman et al. (2011)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Evolutionary strategies:</strong> Stochastic switching</td>
</tr>
<tr>
<td><strong>Comparison method:</strong> Average growth rate &amp; invasion</td>
</tr>
<tr>
<td><strong>Analytical/Numerical:</strong> Both</td>
</tr>
<tr>
<td><strong>Phenotypes:</strong> 2-discrete (x 2-modifier states with recombination)</td>
</tr>
<tr>
<td><strong>Environment:</strong> Periodic (2 states)</td>
</tr>
<tr>
<td><strong>Fitness:</strong> Symmetric</td>
</tr>
<tr>
<td><strong>Selection:</strong> Differential growth rate</td>
</tr>
<tr>
<td><strong>Explicit cost:</strong> None</td>
</tr>
<tr>
<td><strong>Population size:</strong> Infinite (subpopulation frequencies)</td>
</tr>
<tr>
<td><strong>Generations:</strong> Non-overlapping</td>
</tr>
<tr>
<td><strong>Comments:</strong> The authors took Salathé et al. (2009) and Gaál et al. (2010) one step forward including recombination in the model; they observed that, under their model, recombination makes unlikely that a stable non-zero transition rate exists.</td>
</tr>
<tr>
<td>Libby and Rainey (2011)</td>
</tr>
<tr>
<td>-------------------------</td>
</tr>
<tr>
<td><strong>Evolutionary strategies:</strong></td>
</tr>
<tr>
<td><strong>Comparison method:</strong></td>
</tr>
<tr>
<td><strong>Analytical/Numerical:</strong></td>
</tr>
<tr>
<td><strong>Phenotypes:</strong></td>
</tr>
<tr>
<td><strong>Environment:</strong></td>
</tr>
<tr>
<td><strong>Fitness:</strong></td>
</tr>
<tr>
<td><strong>Selection:</strong></td>
</tr>
<tr>
<td><strong>Explicit cost:</strong></td>
</tr>
<tr>
<td><strong>Population size:</strong></td>
</tr>
<tr>
<td><strong>Generations:</strong></td>
</tr>
<tr>
<td><strong>Comments:</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Carja and Feldman (2012)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Evolutionary strategies:</strong></td>
</tr>
<tr>
<td><strong>Comparison method:</strong></td>
</tr>
<tr>
<td><strong>Analytical/Numerical:</strong></td>
</tr>
<tr>
<td><strong>Phenotypes:</strong></td>
</tr>
<tr>
<td><strong>Environment:</strong></td>
</tr>
<tr>
<td><strong>Fitness:</strong></td>
</tr>
<tr>
<td><strong>Selection:</strong></td>
</tr>
<tr>
<td><strong>Explicit cost:</strong></td>
</tr>
<tr>
<td><strong>Population size:</strong></td>
</tr>
<tr>
<td><strong>Generations:</strong></td>
</tr>
<tr>
<td><strong>Comments:</strong></td>
</tr>
<tr>
<td><strong>Kuwahara and Soyer (2012)</strong></td>
</tr>
<tr>
<td>---</td>
</tr>
<tr>
<td><strong>Evolutionary strategies:</strong></td>
</tr>
<tr>
<td><strong>Comparison method:</strong></td>
</tr>
<tr>
<td><strong>Analytical/Numerical:</strong></td>
</tr>
<tr>
<td><strong>Phenotypes:</strong></td>
</tr>
<tr>
<td><strong>Environment:</strong></td>
</tr>
<tr>
<td><strong>Fitness:</strong></td>
</tr>
<tr>
<td><strong>Selection:</strong></td>
</tr>
<tr>
<td><strong>Explicit cost:</strong></td>
</tr>
<tr>
<td><strong>Population size:</strong></td>
</tr>
<tr>
<td><strong>Generations:</strong></td>
</tr>
<tr>
<td><strong>Comments:</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Carja et al. (2013)</strong></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Evolutionary strategies:</strong></td>
<td>Stochastic switching</td>
</tr>
<tr>
<td><strong>Comparison method:</strong></td>
<td>Invasion</td>
</tr>
<tr>
<td><strong>Analytical/Numerical:</strong></td>
<td>Both</td>
</tr>
<tr>
<td><strong>Phenotypes:</strong></td>
<td>4-discrete (x 2-modifier states with recombination)</td>
</tr>
<tr>
<td><strong>Environment:</strong></td>
<td>Periodic &amp; random (n states)</td>
</tr>
<tr>
<td><strong>Fitness:</strong></td>
<td>No explicit assumption</td>
</tr>
<tr>
<td><strong>Selection:</strong></td>
<td>Differential growth rate</td>
</tr>
<tr>
<td><strong>Explicit cost:</strong></td>
<td>None</td>
</tr>
<tr>
<td><strong>Population size:</strong></td>
<td>Infinite (subpopulation frequencies)</td>
</tr>
<tr>
<td><strong>Generations:</strong></td>
<td>Non-overlapping</td>
</tr>
<tr>
<td><strong>Comments:</strong></td>
<td>An extension of Liberman et al. (2011) model; the authors reached similar conclusions.</td>
</tr>
</tbody>
</table>
### Furrow and Feldman (2014)

<table>
<thead>
<tr>
<th>Evolutionary strategies:</th>
<th>Inducible switching; stochastic switching</th>
</tr>
</thead>
<tbody>
<tr>
<td>Comparison method:</td>
<td>Invasion</td>
</tr>
<tr>
<td>Analytical/Numerical:</td>
<td>Simulations</td>
</tr>
<tr>
<td>Phenotypes:</td>
<td>2-discrete (x 2-modifier states with 2 epigenetic states)</td>
</tr>
<tr>
<td>Environment:</td>
<td>Periodic &amp; random (2 states)</td>
</tr>
<tr>
<td>Fitness:</td>
<td>Asymmetric</td>
</tr>
<tr>
<td>Selection:</td>
<td>Differential growth rate</td>
</tr>
<tr>
<td>Explicit cost:</td>
<td>Epigenetic regulation</td>
</tr>
<tr>
<td>Population size:</td>
<td>Infinite (sub-population frequencies)</td>
</tr>
<tr>
<td>Generations:</td>
<td>Non-overlapping</td>
</tr>
</tbody>
</table>

Comments: The authors expanded the classical modifier model (e.g. Salathé et al. (2009)) to consider inducible switching and the associated cost. The authors observed that the environmental fluctuation frequency influences the conditions for evolution of epigenetic regulation (either induced or stochastic switching).

### Carja et al. (2014b)

<table>
<thead>
<tr>
<th>Evolutionary strategies:</th>
<th>Stochastic switching</th>
</tr>
</thead>
<tbody>
<tr>
<td>Comparison method:</td>
<td>Invasion</td>
</tr>
<tr>
<td>Analytical/Numerical:</td>
<td>Both</td>
</tr>
<tr>
<td>Phenotypes:</td>
<td>2-discrete (x 2-modifier states with recombination)</td>
</tr>
<tr>
<td>Environment:</td>
<td>Periodic (2 states)</td>
</tr>
<tr>
<td>Fitness:</td>
<td>Symmetric</td>
</tr>
<tr>
<td>Selection:</td>
<td>Differential growth rate</td>
</tr>
<tr>
<td>Explicit cost:</td>
<td>None</td>
</tr>
<tr>
<td>Population size:</td>
<td>Infinite (subpopulation frequencies)</td>
</tr>
<tr>
<td>Generations:</td>
<td>Non-overlapping</td>
</tr>
</tbody>
</table>

Comments: The authors took Salathé et al. (2009) one step forward including migration in the model, and study the evolution of switching rates in the presence of both spatial and temporal heterogeneity in selection pressures. The authors observed that the evolutionary dynamics of the system are mainly governed by the environmental fluctuation rate.
### Carja et al. (2014a)

<table>
<thead>
<tr>
<th>Evolutionary strategies:</th>
<th>Stochastic switching</th>
</tr>
</thead>
<tbody>
<tr>
<td>Comparison method:</td>
<td>Invasion</td>
</tr>
<tr>
<td>Analytical/Numerical:</td>
<td>Both</td>
</tr>
<tr>
<td>Phenotypes:</td>
<td>4-discrete (x 2-modifier states with recombination)</td>
</tr>
<tr>
<td>Environment:</td>
<td>Periodic &amp; random (2 states)</td>
</tr>
<tr>
<td>Fitness:</td>
<td>Asymmetric</td>
</tr>
<tr>
<td>Selection:</td>
<td>Differential growth rate</td>
</tr>
<tr>
<td>Explicit cost:</td>
<td>None</td>
</tr>
<tr>
<td>Population size:</td>
<td>Infinite (subpopulation frequencies)</td>
</tr>
<tr>
<td>Generations:</td>
<td>Non-overlapping</td>
</tr>
<tr>
<td>Comments:</td>
<td>The authors took Salathé et al. (2009) and Liberman et al. (2011) one step forward including migration in the model, and compare it to the effect of mutation and recombination as sources of phenotypic variation. The authors observed that, under their model, these three essentially different evolutionary forces respond very similar to fluctuating selection.</td>
</tr>
</tbody>
</table>

### Botero et al. (2015)

<table>
<thead>
<tr>
<th>Evolutionary strategies:</th>
<th>Inducible switching; genetic adaptation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Comparison method:</td>
<td>Natural selection</td>
</tr>
<tr>
<td>Analytical/Numerical:</td>
<td>Simulations</td>
</tr>
<tr>
<td>Phenotypes:</td>
<td>Continuous</td>
</tr>
<tr>
<td>Environment:</td>
<td>Periodic (continuous)</td>
</tr>
<tr>
<td>Fitness:</td>
<td>Symmetric</td>
</tr>
<tr>
<td>Selection:</td>
<td>Proportional selection scheme</td>
</tr>
<tr>
<td>Explicit cost:</td>
<td>Phenotypic plasticity</td>
</tr>
<tr>
<td>Population size:</td>
<td>Fixed (discrete; 5000 individuals)</td>
</tr>
<tr>
<td>Generations:</td>
<td>Non-overlapping</td>
</tr>
<tr>
<td>Comments:</td>
<td>The authors used an abstract model which, while simple, can still display plasticity, bet-hedging, and genetic adaptation. Testing multiple environmental variation patterns, they observed that different adaptive responses consistently evolve under different timescales and predictability of the environmental variation.</td>
</tr>
<tr>
<td><strong>Lin et al. (2015)</strong></td>
<td></td>
</tr>
<tr>
<td>----------------------</td>
<td>------------------</td>
</tr>
<tr>
<td><strong>Evolutionary strategies:</strong></td>
<td>Fixed; stochastic switching</td>
</tr>
<tr>
<td><strong>Comparison method:</strong></td>
<td>Frequency</td>
</tr>
<tr>
<td><strong>Analytical/Numerical:</strong></td>
<td>Simulations</td>
</tr>
<tr>
<td><strong>Phenotypes:</strong></td>
<td>2-discrete</td>
</tr>
<tr>
<td><strong>Environment:</strong></td>
<td>Periodic (2 states) + bottlenecks</td>
</tr>
<tr>
<td><strong>Fitness:</strong></td>
<td>Symmetric</td>
</tr>
<tr>
<td><strong>Selection:</strong></td>
<td>Differential growth rate</td>
</tr>
<tr>
<td><strong>Explicit cost:</strong></td>
<td>None</td>
</tr>
<tr>
<td><strong>Population size:</strong></td>
<td>Growing (discrete)</td>
</tr>
<tr>
<td><strong>Generations:</strong></td>
<td>Continuous time</td>
</tr>
<tr>
<td><strong>Comments:</strong></td>
<td>The authors explored the origin of the stochastic transitions in fluctuating environments distinguishing between standing variation and <em>de novo</em> mutations using both an experimental and a mathematical model. The authors concluded that the contribution of each of these mechanisms on the adaptation process depends on the fluctuation timescales.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Belete and Balážsi (2015)</strong></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Evolutionary strategies:</strong></td>
<td>Stochastic switching</td>
</tr>
<tr>
<td><strong>Comparison method:</strong></td>
<td>Average growth rate</td>
</tr>
<tr>
<td><strong>Analytical/Numerical:</strong></td>
<td>Both</td>
</tr>
<tr>
<td><strong>Phenotypes:</strong></td>
<td>2-discrete</td>
</tr>
<tr>
<td><strong>Environment:</strong></td>
<td>Periodic; asymmetric (2 states)</td>
</tr>
<tr>
<td><strong>Fitness:</strong></td>
<td>Asymmetric</td>
</tr>
<tr>
<td><strong>Selection:</strong></td>
<td>Differential growth rate</td>
</tr>
<tr>
<td><strong>Explicit cost:</strong></td>
<td>None</td>
</tr>
<tr>
<td><strong>Population size:</strong></td>
<td>Fixed (discrete; 10000 individuals)</td>
</tr>
<tr>
<td><strong>Generations:</strong></td>
<td>Non-overlapping</td>
</tr>
<tr>
<td><strong>Comments:</strong></td>
<td>The authors explored the stochastic switching rate dependency to the environmental fluctuation frequency in asymmetric environments and fitness as the environmental durations shorten. In this limit, the authors observed that the previously described optimal switching rate matching environmental fluctuation frequency does not always hold.</td>
</tr>
<tr>
<td><strong>Evolutionary strategies:</strong></td>
<td>Genetic adaptation; stochastic switching (bistable genetic circuit)</td>
</tr>
<tr>
<td>-------------------------------</td>
<td>-------------------------------------------------</td>
</tr>
<tr>
<td><strong>Comparison method:</strong></td>
<td>Natural selection</td>
</tr>
<tr>
<td><strong>Analytical/Numerical:</strong></td>
<td>Simulations</td>
</tr>
<tr>
<td><strong>Phenotypes:</strong></td>
<td>Continuous (mechanistic)</td>
</tr>
<tr>
<td><strong>Environment:</strong></td>
<td>Periodic &amp; random (2 states)</td>
</tr>
<tr>
<td><strong>Fitness:</strong></td>
<td>Symmetric (Lorentzian function)</td>
</tr>
<tr>
<td><strong>Selection:</strong></td>
<td>Tournament selection scheme</td>
</tr>
<tr>
<td><strong>Explicit cost:</strong></td>
<td>None</td>
</tr>
<tr>
<td><strong>Population size:</strong></td>
<td>Fixed (discrete; 10000 individuals)</td>
</tr>
<tr>
<td><strong>Generations:</strong></td>
<td>Non-overlapping</td>
</tr>
<tr>
<td><strong>Comments:</strong></td>
<td>Adaptation and survival in fluctuating environ-</td>
</tr>
</tbody>
</table>
...
Appendix D

Evolutionary dynamics of epigenetic switches in fluctuating environments – PIPELINE

D.1 Evolutionary dynamics simulations

To simulate the evolutionary dynamics of a self-regulated gene as described in Chapter 3, three main files are required:

- *EvoDynamics.cpp* – Main instructions
- *GeneNetwork.h* – Library
- *ran1.h* – Random number generator

The code *EvoDynamics.cpp* simulates the evolution of a population with the gene network defined in the *GeneNetwork* class (see *GeneNetwork.h*). The population size (*N*), environmental fluctuation frequency (*nu*), selection pressure (*sT*), mutation frequency (*u*), mutation step size (*M*), gene expression dynamics’ algorithm (alg: 0, deterministic; 1, stochastic/Gillespie), as well as the initial genotype (*k0*; *nH0*; *KD0*) are input arguments defined by the user. The simulation runs for *GMAX*
generations, and the gene expression dynamics for each individual cell are simulated for $T_{MAX}$ time, which corresponds to its life span. The same initial protein number ($A_0$) is assumed for all cells in the population. Some parameters in the gene network model are assumed to be fixed: the basal synthesis activity ($\text{ALPHA}$) and the protein degradation rate ($\text{DEG}$). From the three parameters that define the cell's genotype ($k; n_H; K_D$), which ones are allowed to evolve is determined by the $\text{parE}$ class ($\text{PAR}_E$). The environment fluctuates with a fixed frequency ($\nu$) between two states (0, HIGH; 1, LOW), and the initial environmental state is defined by $E_0$. Each environment is characterized by an optimal protein number ($A_{OPT}$); and the fitness in each environment is calculated using a Lorentzian function centered in the optimal protein number and a width measure $V$. The cells to be cloned in the next generations are selected applying the tournament selection scheme, with a probability $u$ of randomly mutate each of its parameter values. Every generation, the average genotype in the subpopulations is printed in the $AGS$ output file; every cycle, the adaptation strategy per cycle is printed in the $ASC$ output file; and the lineage information per cycle is printed only in the last $C_p$ cycles in these printed in the $LxC$ output file; and finally, the individuals information per epoch for the first nine generation after each environmental change and the last generation of each epoch is printed for the last $E_p$ epochs is printed in the $IxI$ output file.

### D.1.1 Running instructions

1. Compile $EvoDynamics.cpp$ in the presence of the other two files, for example:
   
   ```
   g++ EvoDynamics.cpp -o ED_run.out
   ```

2. Run with the input arguments: population size ($N$), environmental fluctuation frequency ($\nu$), selection pressure (i.e. tournament size, $s_T$), mutation rate ($u$), mutation step size ($M$), algorithm to use (0, if deterministic; 1, if stochastic), and the initial genotype ($k_0; n_H_0; K_D_0$). For example:
In addition of the input arguments, other simulation parameters can be easily modified by the user in the EvoDynamics.cpp file:

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>TMAX</td>
<td>4</td>
<td>Maximum time to simulate each cell during one generation.</td>
</tr>
<tr>
<td>GMAX</td>
<td>10000</td>
<td>Number of generations to simulate.</td>
</tr>
<tr>
<td>Cp</td>
<td>100</td>
<td>Number of cycles to print full lineages’ information.</td>
</tr>
<tr>
<td>Ep</td>
<td>10</td>
<td>Number of epoch to print full individuals’ information (only first 9 &amp; last generation per epoch).</td>
</tr>
<tr>
<td>A0</td>
<td>80</td>
<td>Initial protein concentration.</td>
</tr>
<tr>
<td>E0</td>
<td>1</td>
<td>Initial environmental state (0 if HIGH; 1 if LOW).</td>
</tr>
<tr>
<td>ALPHA</td>
<td>0.25</td>
<td>Basal synthesis activity (a=1 -¿ No feedback).</td>
</tr>
<tr>
<td>DEG</td>
<td>1</td>
<td>Protein degradation constant.</td>
</tr>
<tr>
<td>PAR_E.k</td>
<td>1</td>
<td>Maximum synthesis rate (1 if Evolve; 0 if Fixed).</td>
</tr>
<tr>
<td>PAR_E.nH</td>
<td>1</td>
<td>Hill coefficient (1 if Evolve; 0 if Fixed).</td>
</tr>
<tr>
<td>PAR_E.KD</td>
<td>1</td>
<td>Affinity constant (1 if Evolve; 0 if Fixed).</td>
</tr>
<tr>
<td>AOPT[2]</td>
<td>{80, 20}</td>
<td>Optimal protein number ({A_{opt}^{(H)}, A_{opt}^{(L)}}).</td>
</tr>
<tr>
<td>V</td>
<td>0.2</td>
<td>Width measure for Lorentzian fitness function.</td>
</tr>
<tr>
<td>numRep</td>
<td>10</td>
<td>Total number of replicas to run.</td>
</tr>
</tbody>
</table>

The output files for each replica are:

- Average genotype & subpopulations: File AGS with columns:
1 | Generation $g$
2 | Environment $E$ (0 if HIGH; 1 if LOW)
3 | Bistable fraction $f_B$
4 | $\langle k \rangle_B$
5 | $\langle n_H \rangle_B$
6 | $\langle K_D \rangle_B$
7 | $\langle A \rangle_B$
8 | $\langle w \rangle_B$
9 | $\langle k \rangle_M$
10 | $\langle n_H \rangle_M$
11 | $\langle K_D \rangle_M$
12 | $\langle A \rangle_M$
13 | $\langle w \rangle_M$

These are printed each generation.

• Adaptation strategy per cycle: File ASC with columns:

1 | Generation $g$
2 | PE
3 | IE
4 | BA
5 | GA

These are printed at the end of each environmental cycle. $PE$ corresponds to the fraction of parental lineages with completely bistable lineages and no mutations; $IE$ corresponds to the fraction of parental lineages with completely bistable lineages and exactly one mutation; $BA$ corresponds to the fraction of parental lineages with completely bistable lineages and 2 or more mutations; and $PE$ correspond to the fraction of parental lineages with some monostable genotypes and 2 or more mutations.

• Lineage information per cycle: File $LxC$ with columns:
| 1 | Generation $g$ |
| 2 | Tag of parental lineage |
| 3 | Number of accumulated mutations |
| 4 | Lineage bistability (1 if all genotypes were bistable, 0 otherwise) |
| 5 | Average fitness of the lineage in the cycle $k^{(H)}$ |
| 6 | $n^{(H)}_H$ |
| 7 | $K^{(H)}_D$ |
| 8 | $k^{(L)}$ |
| 9 | $n^{(L)}_H$ |
| 10 | $K^{(L)}_D$ |

These are printed at the end of the last $C_p$ environmental cycles, and $x^{(H)}$, $x^{(L)}$ correspond to the biophysical parameter $x$ at the end of the HIGH and LOW epochs, respectively. Notice that these are not parental lineages, but simply the lineages of the current cycle.

- Individuals information per epoch (first 9 & last generations): File $LxC$ with columns:

| 1 | Generation $g$ |
| 2 | Environment $E$ (0 if HIGH; 1 if LOW) |
| 3 | Cell tag |
| 4 | $k$ |
| 5 | $n_H$ |
| 6 | $K_D$ |
| 7 | $A$ |
| 8 | $w$ |
| 9 | SS1 |
| 10 | SS2 |
| 11 | SS3 |
| 12 | Parent tag |

These are printed at the end of the last $E_p$ epochs, SS# correspond to the phenotype steady state values given the cell’s genotype (if the cell is monostable, SS2 = SS3 = 0). The cell tag simply corresponds to the numbered population, from 0 to N-1, and parent tag tells you the corresponding cell tag of the
parental cell in the previous generation.

In all cases, columns are separated by tabs.

D.1.2 EvoDynamics.cpp

```cpp
#include <iostream>
#include <stdlib.h>
#include <fstream>
#include <stdio.h>
#include <algorithm>
#include <time.h>
#include <cmath>
#include "GeneNetwork.h"

int main(int argc, char *argv[])
{
    #define TMAX 4
    int GMAX = 10000;
    int Cp = 100;
    int Ep = 10;
    // Initial protein concentration:
    int A0 = 80;
    // Initial environmental state:
    int E0 = 1; // 0 – High; 1 – Low
    // Fixed parameters:
    double ALPHA = 0.25;
    double DEG = 1;
    // Evolving parameters (1 – Evolve; 0 – Fixed):
    parE PAR_E;
    PAR_E.k = 1;
    PAR_E.nH = 1;
    PAR_E.KD = 1;
    // Optimal protein number:
    double AOPT[2] = {80, 20};
    double V = 0.2;
    // Replicas:
    int numRep = 10;
    long seeds[10] = {-17,-23,-7,-3,-5,-9,-11,-13,-15,-19};
```
if (argn < 10)
{
    cerr << "Error! Input nine arguments." << endl;
    exit(0);
}

// Input arguments
int N = atoi(args[1]);
double nu = atof(args[2]);
int sT = atoi(args[3]);
double u = atof(args[4]);
double M = atof(args[5]);
bool alg = atoi(args[6]);
double k0 = atof(args[7]);
double nH0 = atof(args[8]);
double KD0 = atof(args[9]);

for (int iS = 0; iS < numRep; iS++)
{
    // OUTPUT FILES
    char myFileName[255];
    // Average genotype & subpopulations:
    sprintf(myFileName, "AGS%d.dat", iS);
    ofstream AGS(myFileName, ios::out);
    // Adaptation strategy x cycle:
    sprintf(myFileName, "ASC%d.dat", iS);
    ofstream ASC(myFileName, ios::out);
    // Lineage information x cycle:
    sprintf(myFileName, "LxC%d.dat", iS);
    ofstream LxC(myFileName, ios::out);
    // Individuals information per epoch
    // (first 9 & last generations):
    sprintf(myFileName, "lxE%d.dat", iS);
    ofstream lxE(myFileName, ios::out);

    long seed = seeds[iS]; // Initialize random seed.
    int E = 1 - E0; // Initialize environment.

    // Create the cell population:
    GeneNetwork myPop[N];
    PopAverage myPopB;
    PopAverage myPopM;
// Assign user-defined-parameters values:
myPop[0] . ic (A0, k0, ALPHA, nH0, KD0, DEG);
myPop[0] . LTag = 0;
for (int i = 1; i < N; i++)
{
    myPop[i] = myPop[0];
    myPop[i] . LTag = i;
}

// Simulate generations:
for (int iG = 1; iG < GMAX; iG++)
{
    // Update environmental state each epoch:
    if (iG % int (1 / nu) == 1)
    {
        E = (E + 1) % 2;
    }

    myPopB . reset();
    myPopM . reset();

    // Loop over the population:
    for (int i = 0; i < N; i++)
    {
        // Find the solutions of the system (SS)
        // only if a mutation has occurred:
        if (myPop[i] . mut)
        {
            myPop[i] . findSS();
        }

        // Gillespie Algorithm:
        if (alg)
        {
            while (myPop[i] . time < TMAX)
            {
                myPop[i] . updateTime(&seed);
                myPop[i] . updateSystem(&seed);
            }
        }

        // Deterministic Algorithm (assuming steady state
        // is reached in the life span):
    }
else
{
   // If the cell is bistable & the cell was previously
   // in the "high" state, stay in this state:
   if((myPop[i].SS[2] > 0) &&
      (myPop[i].SS[1] < myPop[i].A))
   {
      myPop[i].A = myPop[i].SS[2];
   }
   else
   {
      myPop[i].A = myPop[i].SS[0];
   }
}

// Calculate the fitness of cell [i]:
myPop[i].calculateFitness(E, AOPT, V);
myPop[i].LFit += (myPop[i].w*(nu/2));

// MONOSTABLE
if(myPop[i].SS[2] == 0)
{
   myPop[i].LBis = 0;
   myPopM.k += myPop[i].k;
   myPopM.nH += myPop[i].nH;
   myPopM.KD += myPop[i].KD;
   myPopM.A += myPop[i].A;
   myPopM.w += myPop[i].w;
   myPopM.f++;  
}
// BISTABLE
else
{
   myPopB.k += myPop[i].k;
   myPopB.nH += myPop[i].nH;
   myPopB.KD += myPop[i].KD;
   myPopB.A += myPop[i].A;
   myPopB.w += myPop[i].w;
   myPopB.f++;  
}

// PRINTING – Individuals information per epoch
   // (first 9 & last generations):
   //
   //
\textbf{if} ((iG \geq (GMAX - (Ep/nu))) && (iG\%int(1/nu) < 10))
{
  IxE << iG << \\
  IxE << myPop[i].k << \\
  IxE << myPop[i].KD << \\
  IxE << myPop[i].w << \\
  IxE << myPop[i].SS[0];
  IxE << ''t'' << myPop[i].SS[1] << ''t'';
  IxE << myPop[i].SS[2] << ''t'';
  IxE << myPop[i].parent << endl;
}

// PRINTING – Average genotype & subpopulations:
myPopB. normalize(N);
myPopM. normalize(N);
AGS << iG << ''t'' << E << ''t'' << myPopB.f << ''t'';
AGS << myPopB.k << ''t'' << myPopB.nH << ''t'';
AGS << myPopB.KD << ''t'' << myPopB.A << ''t'';
AGS << myPopB.w << ''t'' << myPopM.k << ''t'';
AGS << myPopM.nH << ''t'' << myPopM.KD << ''t'';
AGS << myPopM.A << ''t'' << myPopM.w << endl;

// Each epoch, record genotype per lineage:
if (iG\%int(1/nu) == 0)
{
  for (int i = 0; i < N; i++)
  {
    myPop[i].kE[E] = myPop[i].k;
    myPop[i].nHE[E] = myPop[i].nH;
    myPop[i].KDE[E] = myPop[i].KD;
  }
}

// Each cycle:
if (iG\%int(2/nu) == 0)
{
  double AS[5] = {0, 0, 0, 0, 0};
  for (int i = 0; i < N; i++)
  {
    myPop[i].adaptS();
    AS[myPop[i].ASp]++;
  }
}
if (iG > (GMAX - (Cp * 2 / nu)))
{
    // PRINTING — Lineage information x cycle:
    LxC << iG << 't' << myPop[i].LTag << 't';
    LxC << myPop[i].LMut << 't' << myPop[i].LBis << 't';
    LxC << myPop[i].LFit << 't' << myPop[i].kE[0] << 't';
    LxC << myPop[i].nHE[0] << 't' << myPop[i].KDE[0];
    LxC << 't' << myPop[i].kE[1] << 't';
    LxC << myPop[i].nHE[1] << 't';
    LxC << myPop[i].KDE[1] << endl;
}

// Reset lineage’s statistics:
myPop[i].LMut = 0;
myPop[i].LBis = 1;
myPop[i].LFit = 0;
myPop[i].LTag = i;

// PRINTING — Adaptation strategy x cycle:
ASC << iG << 't';
ASC << AS[1] / N << 't';
ASC << AS[2] / N << 't';
ASC << AS[3] / N << 't';

// Apply selection and evolve parameters:
selectionTournament(myPop, N, sT, u, M, &seed, PAR_E);
}

// Close files
AGS.close();
ASC.close();
LxC.close();
IxE.close();
}
return 0;
}

D.1.3 GeneNetwork.h

// Libraries:
#include <stdio.h>
```cpp
#include <cmath>
#include <stdlib.h>
#include <time.h>
#include <iostream>
#include <algorithm>
#include <vector>
#include "ran1.h"

using namespace std;

class parE
{
  public:
    bool k, a, nH, KD, d;

    parE()
    {
      k = 0;
      a = 0;
      nH = 0;
      KD = 0;
      d = 0;
    }
};

class PopAverage
{
  public:
    double k, nH, KD;
    double A, w, f;

  void reset()
  {
    k = 0;
    nH = 0;
    KD = 0;
    A = 0;
    w = 0;
    f = 0;
  }

  void normalize(int N)
  {
```
\begin{verbatim}
k = k/f;
nH = nH/f;
KD = KD/f;
A = A/f;
w = w/f;
f = f/N;

class GeneNetwork
{
public:
    // Genotype:
    double k, nH, KD;
    double a, d;
    // Phenotype:
    double A;
    double SS[3];
    // Evolutionary metrics:
    double w;
    int parent; // Parent cell number
    bool mut; // Flag: 0 if parental genotype; 1 if mutant
    // Lineage metrics per cycle:
    int LTag; // Tag for parental lineage
    int LMut; // Number of accumulated mutations per cycle
    bool LBis; // Flag for bistable lineages.
    double LFIt; // Average fitness of the lineage.
    // Genotype at the end of each epoch [HIGH,LOW]:
    double kE[2];
    double nHE[2];
    double KDE[2];
    // Flag for the adaptation strategy used by the lineage
    // (0 = Ma; 1 = PE; 2 = IE; 3 = BA; 4 = GA):
    int AS;
    int ASp; // AS in parental lineage.
    // Gene dynamics variables:
    double r[2], rT, time;
}

GeneNetwork()
{
    w = 0;
    parent = 0;
    mut = 0;
}\end{verbatim}
LMut = 0;
LBis = 1;
LFit = 0;
AS = 0;

time = 0;
}

// Assign initial genotype & phenotype:
void ic(double A0, double k0, double a0, double nH0,
        double KD0, double d0)
{
    A = A0;    // Protein number
    k = k0;    // Maximum synthesis rate
    a = a0;    // Basal synthesis activity
    nH = nH0; // Hill coefficient
    KD = KD0; // Affinity constant
    d = d0;   // Protein degradation constant
    findSS(); // Calculate steady states (SS)
}

// Find the steady state solutions:
void findSS()
{
    double Ai = 0;
    int ss = 0;
    double FA0, FA;

    // Initialize SS:
    SS[0] = 0;
    SS[1] = 0;
    SS[2] = 0;

    // Initial value of the equilibrium equation:
    FA0 = Ai - (k/d)*(a+((1-a) *
                        *(pow(Ai,nH)/(pow(KD,nH)+pow(Ai,nH)))));

    for (Ai=0.1; Ai <= 1000; Ai=Ai+0.1)
    {
        // Calculate equilibrium equation:
        FA = Ai - (k/d)*(a+((1-a) *
                         *(pow(Ai,nH)/(pow(KD,nH)+pow(Ai,nH)))));

        // other code...
    }

    // other code...
}

// other code...
// If there is a change of sign in the equilibrium equation, there should be a steady state:
if (FA*FA0 <= 0)
{
    SS[ss] = Ai;
    ss += 1;
}

// It stops if the system already have 3 steady states:
if (ss == 3)
{
    break;
}

// Make the actual value the previous value and iterate:
FA0 = FA;
}

// Update reaction propensities:
void updateR()
{
    // Synthesis reaction:
    r[0] = k*(a+((1-a)*(pow(A,nH)/(pow(KD,nH)+pow(A,nH)))));
    // Degradation reaction:
    r[1] = d*A;
    rT = r[0]+r[1]; // Total propensity
}

// Perform the Gillespie algorithm:
void updateTime(long *seed)
{
    // Random number between 0 and 1:
    double rand1 = (double)ran1(seed);
    // Time between reactions:
    double re_time;
    // Update reaction propensities:
    updateR();
    // Time for the next reaction:
    re_time = -log(1-rand1)/rT;
    // Actualize global time:
    time += re_time;
}

void updateSystem(long *seed)
{
// Random number between 0 and 1:
double rand2 = (double)ran1(seed);
if (rand2 < r[0]/rT)
{
    A++; // Synthesis reaction
}
elif (rand2 <= (r[0]+r[1])/rT)
{
    A--; // Protein degradation
}
else
{
    cerr << "Runtime Error: rT not real." << endl;
    exit(0);
}

// Evaluate cell’s fitness (Lorentzian function):
void calculateFitness(int myE, double Aopt[2], double v)
{
    w = 1/(1+((A-Aopt[myE])*(A-Aopt[myE])/(v*Aopt[myE])));
}

// Random mutation in the genotype:
void mutate(long *seed, double M, parE PAR_E)
{
    #define PI 3.14159265
    // Random number between 0 and 1:
    double ratioQ = (double)ran1(seed);
    // Random numbers between 0 and 2*PI:
    double thetaQ = (double)ran1(seed)*2*PI;
    double phiQ = (double)ran1(seed)*2*PI;

    if(PAR_E.k)
    {
        k *= pow(M, ratioQ*sin(thetaQ)*cos(phiQ));
        if(k>1000) { k=1000; }
        if(k<0.01) { k=0.01; }
    }
    if(PAR_E.nH)
    {
        nH *= pow(M, ratioQ*sin(thetaQ)*sin(phiQ));
        if(nH>16) { nH=16; }
    }
if (nH<0.01) { nH=0.01; }
}
if (PAR_E.KD)
{
    KD *= pow(M, ratioQ*cos(thetaQ));
    if (KD>120) { KD = 120; }
    if (KD<0.01) { KD=0.01; }
}
mut = 1; // Update mutation flag.
LMut++; // Add to mutations count.

// Define lineage adaptation strategy:
void adaptS()
{
    ASp = AS;
    if (LBis==1)
    {
        if (LMut==0)
        {
            AS = 1; // Pure epigenetic (PE): Bis & No mut
        }
        else if (LMut==1)
        {
            AS = 2; // Impure epigenetic (IE): Bis & 1 mut
        }
        else
        {
            AS = 3; // Bistable adaptation (BA): Bis & >1 mut
        }
    }
    else
    {
        if (LMut>1)
        {
            AS = 4; // Genetic adaptation (GA): No Bis & >1 mut
        }
        else
        {
            AS = 0; // Maladapted (Ma): No Bis & <2 mut
        }
    }
}
void selectionTournament(GeneNetwork *myPop, int N, int t,
    double u, double M, long *seed,
    parE PAR_E)
{
    GeneNetwork newPop[N];  // The new population to be defined.

    // Fill in the next generation:
    int iN = 0;
    int rand1;
    while (iN < N)
    {
        // Stochastically choose the tournament cells and
        // replicate the cell with higher fitness in the group:
        int indexT[t];
        indexT[0] = int(N*(double)ran1(seed));
        int maxT = indexT[0];
        int j = 1;
        while (j < t)
        {
            rand1 = int(N*(double)ran1(seed));
            for (int i=0; i < j; i++)
            {
                if (rand1 == indexT[i])
                {
                    rand1 = -1;
                    break;
                }
            }
            if (rand1 > 0)
            {
                indexT[j] = rand1;
                if (myPop[maxT].w < myPop[indexT[j]].w)
                {
                    maxT = indexT[j];
                }
                j++;
            }
        }
        // The new cell is equal to the old cell selected:
        newPop[iN] = myPop[maxT];
    }
}
newPop[iN].parent = maxT; // Parent cell number
iN++;
}

// Refill the population matrix & allow random mutation:
for (int i = 0; i < N; i++)
{
    myPop[i] = newPop[i];
    myPop[i].time = 0; // Reset the time to zero.
    myPop[i].mut = 0; // Flag for parental genotype.
    if ((double)ran1(seed) <= u)
    {
        myPop[i].mutate(seed, M, PAR_E);
    }
}

D.1.4 ran1.h

The “Minimal random number generator of Park and Miller with Bays-Durham shuffle and added safeguards” from Numerical Recipes in C++ (Press et al., 2002) was used to generate random numbers. A negative integer is required to initialize (i.e. seed).
Bibliography


208


210


211


Biography

Mariana Gomez Schiavon was born on September 7th, 1986, in Puebla, Mexico. She obtained her B.S. degree in Genome Sciences from the National Autonomous University of Mexico (UNAM) in 2009, and her M.S. degree in Biomedical Engineering and Physics from Center for Research and Advanced Studies (CINVESTAV), Campus Monterrey, in 2011. In CINVESTAV, Mariana studied gene circuit dynamics and stochasticity under the advisory of Prof. Moiss Santilln. In August 2011, she started her PhD in Computational Biology and Bioinformatics in Duke University, where she joined the laboratory of Prof. Nicolas Buchler. Here, Mariana integrated her background on gene circuit dynamics and the role of a biochemical noise, with evolutionary theory. She is interested in understanding how emergent properties of gene networks appear, spread and persist through evolution.