

GRiNS: A Python Library for Simulating Gene Regulatory Network Dynamics

Pradyumna Harlapur¹, Harshavardhan B V², and Mohit Kumar Jolly^{1,*}

¹Department of Bioengineering, Indian Institute of Science, Bengaluru, Karnataka, India - 560012

²IISc Mathematics Initiative, Indian Institute of Science, Bengaluru, Karnataka, India - 560012

*Corresponding Author, Email: mkjolly@iisc.ac.in

Abstract

The emergent dynamics of complex gene regulatory networks govern various cellular processes. However, understanding these dynamics is challenging due to the difficulty of parameterizing the computational models for these networks, especially as the network size increases. Here, we introduce a simulation library, Gene Regulatory Interaction Network Simulator (GRiNS), to address these challenges. GRiNS integrates popular parameter-agnostic simulation frameworks, RACIPE and Boolean Ising formalism, into a single Python library capable of leveraging GPU acceleration for efficient and scalable simulations. GRiNS extends the ordinary differential equations (ODE) based RACIPE framework with a more modular design, allowing users to choose parameters, initial conditions, and time-series outputs for greater customisability and accuracy in simulations. For large networks, where ODE-based simulation formalisms do not scale well, GRiNS implements Boolean Ising formalism, providing a simplified, coarse-grained alternative, significantly reducing the computational cost while capturing key dynamical behaviours of large regulatory networks. The documentation and installation instructions for GRiNS can be found at <https://moltenecdysone09.github.io/GRiNS/>.

Keywords:

Gene Regulatory Networks, Network Dynamics, Parameter-Agnostic Simulation, Ising Boolean Formalism, Random Circuit Perturbation (RACIPE), Systems Biology

1 Introduction

One of the core goals of systems biology is to understand and model complex interactions in a biological system to understand the mechanistic underpinnings of various cellular processes. The advent of high-throughput technologies capable of capturing the molecular profiles of cells has enabled us to infer causal relations between molecules governing various biological processes [1, 2, 3]. These causal interactions between genes are typically represented using gene regulatory networks (GRNs). GRNs are network representations of the causal interactions between genes that govern their expression levels and functional activity [4]. GRNs are crucial for understanding the mechanistic details governing complex biological processes that give rise to specific expression patterns and

cellular phenotypes [5, 6, 7, 8]. As representations of the interactions between genes, GRNs provide a framework to decipher the regulatory logic and identify key interactions that determine a particular biological outcome. There are two main approaches to constructing GRNs [9]. The bottom-up approach employs detailed experiments to verify the interactions between genes to build these GRNs, with experiments to tease out causal relations between genes. However, while being precise, this approach is cumbersome because, with the increasing network size, it becomes complicated to accurately determine all the interactions between the genes. On the other hand, the top-down approach leverages high-throughput technologies and the vast amounts of data they generate to infer the interactions between genes. Various algorithms have been developed that utilize and integrate data from different single-cell molecular profiling techniques to infer gene regulatory networks [1, 2, 3, 10]. However, unlike the bottom-up approach, such inferred GRNs often have limited predictive potential due to missing or inaccurate regulatory links [11].

When it comes to understanding the dynamics of such predicted GRNs, most of these methods suffer from a common challenge of being unable to accurately predict the interaction functions and their parameters [11, 12]. Given the uncertainty in data due to biological noise and technical limitations, and considering the sizes of the large networks that govern these interactions, it becomes challenging to parameterize these networks accurately [13, 14]. Having simulation frameworks that are parameter agnostic and focus on the dynamics and steady states of a given network structure becomes important in such cases. These approaches do not depend on specific parameter sets to explain the behavior of GRNs. An advantage of such approaches is that, even in the absence of parameters, having qualitative estimates of the general ranges of the parameters is sufficient for getting an idea about the dynamics of the network in the parameter range.

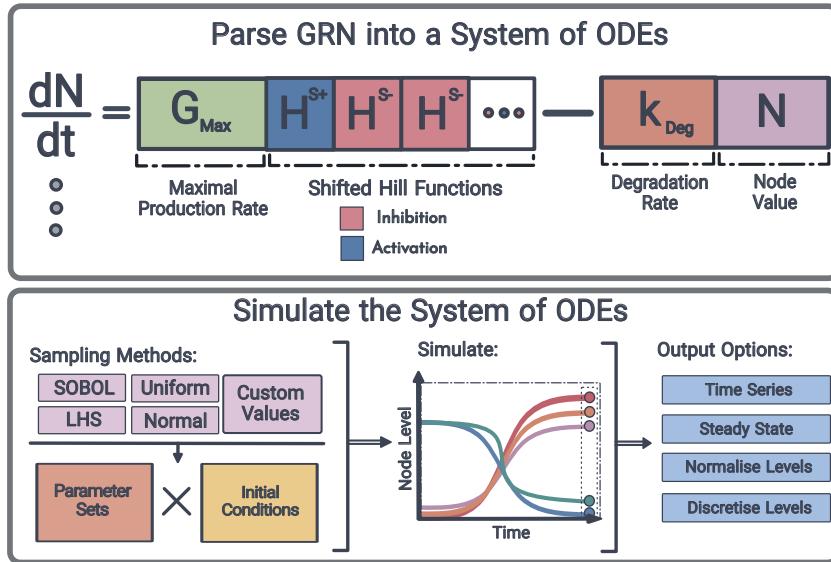
RAndom CIrcuit PErturbation (RACIPE) is one such framework that tries to identify the possible phenotypic space of a network given its topology [15, 16]. It samples parameters over predefined ranges and simulates them over multiple initial conditions. The tool can capture a network’s steady states by randomly sampling the parameters and simulating them over multiple initial conditions. This approach helps one to understand the possible types of dynamics and steady states a network can show, even when the mathematical model’s precise kinetic parameters are absent. Additionally, due to the random sampling of parameters and initial conditions, the RACIPE tool mimics the wide range of variability observed in biology, giving us a more realistic view of the network’s behavior compared to when we use precise parameters, which, more often than not, are very context-specific and may not represent the actual behavior of the network in a different scenario. As RACIPE automates the entire pipeline of model building and its subsequent simulation according to the inputs provided by the user, it proves to be a suitable tool for analyzing the dynamics of GRNs. It has been used to model and explain various cellular processes, including cell fate decisions, phenotypic heterogeneity, and transitions such as epithelial–mesenchymal plasticity across diverse biological contexts [17, 18, 6].

Another such parameter-agnostic method is the Ising boolean simulation framework [19, 20]. This framework represents each gene as a binary variable—active or inactive—depending on the cumulative influence of the incoming active links. Since each update step is based on matrix multiplication, this method, although very crude, is suitable for simulating large networks where the system of ODEs is too large and, hence, is too slow to be practical to be simulated over thousands of parameters and initial conditions through RACIPE. It has been applied to analyze gene networks and their state transitions with

minimal parameter dependence [19, 21, 22].

Because both these methods depend heavily on matrix-based operations, they are ideal candidates for GPU implementation. GPUs are suitable for matrix-based operations because they can parallelize the computation, giving dramatic speed-ups compared to the same computations run on a CPU. The Python library *Jax*'s efficient array-oriented numerical computation functions provide an excellent foundation for our simulation library [23]. RACIPE, in particular, requires GPU-based differential equation solver implementations provided by the *diffraex* [24] library, which is built on top of *Jax*. Utilizing these, we introduce a simulation library for GRNs that exploits the speed of GPUs over CPUs for matrix operations, enabling rapid and scalable simulations of GRNs (Fig. 1). Our library's modular design, written in an easily accessible language like Python, offers more options for the users to tweak the simulations according to their needs.

• RAndom Circuit PErurbation (RACIPE)



• Ising Boolean Formalism

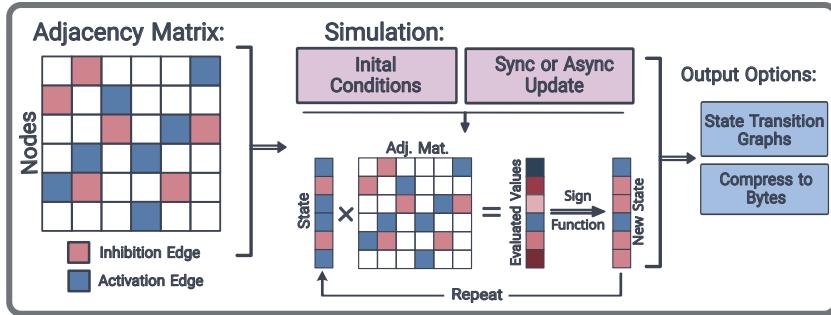


Figure 1: Overview of the simulation frameworks in GRiNS. GRiNS includes implementations of Random Circuit Perturbation (RACIPE) for continuous ODE-based modeling and Ising Boolean formalism for discrete-state simulations.

2 Methods

2.1 RAndom CIrcuit PErturbation (RACIPE)

RACIPE is a GRN modeling framework designed to sample the steady-state repertoire of a GRN purely based on its topological structure [16]. It does this by generating a system of coupled ordinary differential equations (ODEs) to represent the interactions of the genes and then simulating these equations over multiple randomly sampled parameter sets and initial conditions. By simulating a parameterized system of ODE over the initial conditions, RACIPE can uncover the possible set of steady states. When this process is carried out on all the parameter sets, RACIPE can map out the possible phenotypic outcomes of a network even in the absence of specific parameter sets.

In the following sections, we describe the methodology of RACIPE and how it constructs the differential equation models from network topologies and systematically samples parameters within biologically relevant ranges. We will then describe the simulation methodology of the parameterized ODEs over the initial conditions and the subsequent analysis, which can be done to identify robust, steady states to uncover the dynamic landscape of gene regulatory networks.

2.1.1 Parsing GRNs to Construct System of ODEs

RACIPE models a signed and directed GRN as a system of coupled ODEs. For a gene, T, in the GRN, the ODE describing the change in its expression value as a function of the input nodes is given by:

$$\frac{dT}{dt} = G_T * \prod_i \frac{H^S(P_i, P_{iT}^0, n_{PiT}, \lambda_{PiT})}{\lambda_{PiT}} * \prod_j H^S(N_j, N_{iT}^0, n_{NjT}, \lambda_{NjT}) - k_T * T \quad (1)$$

Where,

$$H^S(B, B_A^0, n_{BA}, \lambda_{BA}) = \frac{B_A^{0n_{BA}}}{B_A^{0n_{BA}} + B^{n_{BA}}} + \lambda_{BA} * \frac{B^{n_{BA}}}{B_A^{0n_{BA}} + B^{n_{BA}}} \quad (2)$$

G_T in Equation (1) refers to the maximal expression value of the node T, H^s is a modified form of Hill's equation called the Shifted Hill's equation, as given in Equation (2). Each Shifted Hill's Equation (2) represents the effect of an upstream incoming node on T. It has the threshold parameter B^0 , n_{BA} is the hill's coefficient, and λ_{BA} is the fold change parameter representing the fold change in the expression of A brought about by the effect of the edge from node B. P_i , and N_j refers to the values of the input activating and inhibiting nodes respectively. All the hill's terms of the incoming edges are multiplied with the maximal expression value G_T to get a scaled value of T's expression, the production term of the ODE. Additionally, k_T is the degradation rate of the node and is multiplied by the value of T to get the degradation term of the equation.

The current parser supports only signed and directed GRNs, i.e., it only recognizes activation and inhibition links when constructing the system of ODEs. Once generated, the ODE system is written into a Python file formatted for compatibility with the Diffrazx simulation library. Users can modify these functions to incorporate custom dynamics, as the subsequent simulation steps do not depend on the original RACIPE ODE structure given above.

2.1.2 Sampling Parameters and Initial Conditions

Since RACIPE’s primary object is to capture all possible steady states of a given GRN, the parameter sampling strategy must eliminate any biases that may skew the result. The number of parameters to be sampled for a network with N nodes and E edges is $2N + 3E$. The term $2N$ represents the maximal production and degradation rate parameters that must be sampled. The $3E$ term represents the threshold, the hill’s coefficient, and the fold change parameters that need to be sampled for each of the edges present in the network. Table 1 gives the default ranges of the parameter values over which the sampling is done. Th

Parameters	Minimum	Maximum
Production Rate (G)	1	100
Degradation Rate (k)	0.1	1
Fold Change (Inhibition λ)	0.01	1
Fold Change (Activation λ)	1	100
Hill Coefficient (n)	1	6
Threshold	The ranges depend on the in-degree - half functional rule	

Table 1: Default parameter ranges used by RACIPE.

The inhibition fold change parameter is sampled in the inverse range, i.e., between the inverse of the maximum value and the inverse of the minimum value. Following this, the inverse of all the sampled values is taken. This process shifts the mean of the distribution from 0.5 to 0.02 (in the case of default ranges), which ensures that both weak and strong inhibitory fold change parameters are sampled in a balanced manner.

Another aspect to consider to eliminate biased parameters of nodes is the threshold values assigned to the edges. For nodes with incoming edges, the threshold values need to be sampled to be within the minimal and maximal steady-state expression ranges of the nodes from which they originate. Otherwise, a deviation from this would mean that an edge could be always active or always inactive, biasing the simulations. The half-functional rule is employed to correct this bias, and it ensures that the threshold values are chosen between 0.02 and 1.98 times the median steady-state expression level of the upstream node. Suppose other genes regulate the upstream node itself. In that case, its steady-state expression distribution is determined by simulating its steady-state levels based on the regulatory parameters of its incoming nodes. The median expression of the upstream node is then used to define the appropriate range for threshold sampling. This ensures that the threshold values are chosen so that each regulator has a probability of being active at roughly 50% across all the simulations.

The values are sampled from the respective node’s minimum and maximum expression values for the initial conditions. Our framework supports multiple sampling methods—including Sobol (default), Latin hypercube, uniform, log-uniform, normal, and log-normal distributions—and even allows mixing different distributions for different parameters or initial conditions [25]. This flexibility lets users target specific regions of parameters or the initial condition space. A parameter range file is also generated for easy customization, on which the parameters can be regenerated for subsequent simulations. Similarly, initial conditions can also be regenerated based on the updated ranges.

2.1.3 Simulating the ODE system and processing the results

Once the parameters and initial conditions are generated after the GRN is parsed to create the ODE file, the library provides functions for the simulation of the ODE system for all combinations of parameter sets and initial condition values. A function is also provided to simulate the ODE system for a single combination of parameters and initial conditions. The simulations are usually done using the *vmap* function of the *Jax* library, and the user can control the batch size for cases where the VRAM is insufficient to run large-scale simulations. The user can control the tolerances (relative and absolute), the start time point (by default is zero as we would be dealing with initial conditions), and the end time point of the simulations, after which the simulation will terminate even though the steady-state may not have been reached. By default, the simulations terminate once the steady state condition is reached, which is determined by the tolerance values. The resulting data—comprising node values, termination times, a steady-state indicator, and identifiers for each parameter–initial condition pair—is stored in a data frame as a Parquet file on the disk.

Setting the end time point sufficiently high is important to allow most combinations to reach a steady state, which the steady state flag column of the solution data frame can track. The implementation also supports recording gene expression levels at custom time points for time-series analyses. However, we recommend first determining the steady-state range and then deciding on the optimal spacing of the time points and their range to capture the relevant dynamics.

Since parameters can vary across each simulation instance, a normalization method is necessary to compare steady-state values. Our library provides functions that normalize node expression by the maximal expression-to-degradation ratio (G/k) – the highest possible expression level – resulting in outputs scaled between 0 and 1, with 1 being the maximal expression of the node. Additionally, nodes can be discretized into levels based on the global distribution of normalized G/k values, offering a standardized approach to processing simulation outputs. These functions are compatible with both the steady-state and time-series simulation results.

2.2 Boolean Ising Formalism

Although RACIPE formalism is an excellent framework for understanding the possible dynamics emergent from a GRN, computational cost becomes prohibitive for large networks due to the increasing number of parameters and longer simulation times. Boolean Ising formalism, also referred to as threshold boolean formalism, provides a simpler, albeit coarse-grained approach for simulating large networks [22, 20, 4]. In this approach, the state of the variable is represented by a discrete variable indicating the on or off state of the gene. The system starts with an initial condition (s_0), and the subsequent state (s_{t+1}) is obtained by the previous state’s (s_t) matrix product with the network’s adjacency matrix (A), where -1 represents inhibitory interactions, 1 represents activating interactions, and 0 indicates no connection. Once the state is multiplied with the adjacency matrix, the resultant vector is then converted to 1s and -1s (or 0s depending on the flip values provided) according to the rule given below:

$$s_i(t+1) = \begin{cases} 1, & \sum_j J_{ij} s_j(t) > 0 \\ s_i(t), & \sum_j J_{ij} s_j(t) = 0 \\ -1, & \sum_j J_{ij} s_j(t) < 0 \end{cases} \quad (3)$$

The system can be updated using either a synchronous update, where all the nodes are updated simultaneously, or an asynchronous update, where one randomly chosen node is updated at each step. The user can set the number of simulation steps, define initial conditions (or use randomly generated ones by default), and specify a custom update order for asynchronous mode. The simulation results are stored in a dataframe with columns for the binary node states (0 or 1), the simulation step number, and the corresponding initial condition number. Optionally, the node values at each step can be stored as bytes to conserve disk space and later unpacked during analysis.

3 Case Studies

3.1 Comparison of the steady state dynamics of Toggle Switch and a Toggle Switch with Self-activation

To demonstrate the utility of our simulation framework, we present a comparative analysis of the toggle switch (TS) (Figure 2A, i) and its variant with self-activation on both nodes (TSSA) (Figure 2A, ii). The TS motif is commonly found in gene regulatory networks and is known for exhibiting bistability, enabling cells to maintain expression states that let them commit to distinct fates [26, 27, 28]. The mutual inhibition between the two genes results in antagonistic expression profiles—an essential property for cell fate commitment—as it prevents promiscuous co-expression of the genes belonging to both the cell types [29].

We use the RACIPE implementation in the GRiNS library to simulate these motifs and contrast the effect of adding self-activating loops to the TS motif in terms of steady states, the distribution of multistable states, and the normalization and analysis of steady-state data across different parameter sets and initial conditions. We simulated the motifs over 10,000 parameter sets and 1,000 initial conditions, run in triplicate. Sobol sampling was used to generate both the parameters and the initial conditions. During the parameter and initial condition generation step, the function also parses the GRN topology file to generate a Python file containing the ODE representation of the GRN. The ODE system follows the structure of the RACIPE formalism laid out in the Methods section. Once generated, the ODE system file is saved in a user-specified folder named after the input topology file. The user can then also manipulate the system ODE file, providing flexibility to introduce custom functions in addition to the default shifted Hill functions used by RACIPE. In such cases, however, the generated parameters may not be compatible with simulations, and a custom parameter file would be required. The rationale for running simulations in triplicate is to ensure sufficient sampling of parameters and initial conditions from a given topology file. Since the RACIPE framework is parameter-agnostic, with its primary purpose being the identification of steady-states, their types, and distributions, it is important to assess the granularity of parameter space sampling. Too coarse a sampling could miss rare but important states; too dense a sam-

pling increases computation time per replicate. We deemed the number of parameters and initial conditions sufficient for the current use case.

After parameter and initial condition generation, the Python ODE file is loaded and simulated in parallel using the diffraex library’s Tsit5 ODE solver. This GPU-compatible solver significantly speeds up simulations, particularly suitable for the RACIPE framework, where the same ODE system is parameterized and simulated over many independent initial conditions. As each simulation is independent of the others, this presents an opportunity to parallelize the simulations using the GPUs to speed up the simulations.

Simulation results are reported as a parquet file containing the parameter and initial condition indices, a steady-state flag (one if no steady state is reached within the user-defined time step, zero otherwise), and the steady-state values of the nodes. Because parameter values are randomly sampled, directly comparing raw steady-state values across topologies or parameter sets is not meaningful. To address this, we apply G/k normalization to scale steady-state values between 0 and 1, where 1 indicates the maximal possible expression level for a gene under the given parameter set. For both TS and TSSA motifs, the heatmaps show that most steady states involve one node expressing close to 1 (high expression) and the other close to 0 (low expression), consistent with mutually exclusive expression (Figure 2B, i; Figure 2C, i).

However, since these normalized values are still continuous, it becomes difficult to analyze and compare the state spaces of large or multiple GRNs. To address this, a function in our library allows coarse-grain expression levels into discrete “levels,” simplifying comparisons across parameter sets and networks. This approach simplifies the further use of complex dimensionality reduction or clustering algorithms, making it scalable to larger GRNs. After the coarse-graining step, the steady states of TS and TSSA reduce to binary values (0s and 1s) (Figure 2B, ii; Figure 2C, ii). The binary nature of the levels reveals that both TS and TSSA motifs primarily exhibit two expression levels per node.

The highly bimodal nature of the motif nodes is evident in their G/k normalized expression values, too, which are either very high, close to 1, or very low, close to 0. Based on the levels of the nodes for a particular parameter and initial condition set, we then assign it a state composed of the concatenated levels of the nodes. We then examined the distribution of these coarse-grained steady states for multistability. The multistability of a parameter refers to the number of unique steady states it produces when simulated across all initial conditions. Most of the sampled TS parameters were monostable, while TSSA shows an increased number of bistable and a small number of tristable parameters (Figure 2D). Next, when we looked at the distribution of the states themselves, consistent with expectations from the heatmaps, both motifs most frequently produce the 10 and 01 states. Although all four binary combinations (00, 01, 10, 11) are observed, TSSA exhibits a higher frequency of the 00 state than TS, which shows both 00 and 11, albeit rarely (Figure 2E).

Continuing with the multistability analysis of the parameter sets, we observed that TSSA frequently exhibits bistability between 01 and 10 (Figure 2F). At the same time, TS mainly shows monostable 01 or 10 states, with bistable 01–10 states occurring less frequently. TSSA produces tristable states that consist of the 00, 01, and 10 states, which explains the higher occurrence of tristability in TSSA compared to TS. Going beyond the mere presence of multistability, one can also examine the relative frequencies of the coarse-grained steady states for a particular parameter set across multiple initial conditions. To illustrate this, we randomly selected three bistable parameter sets that differ in the frequencies of the 01 and 10 states they produce. The plots of the initial

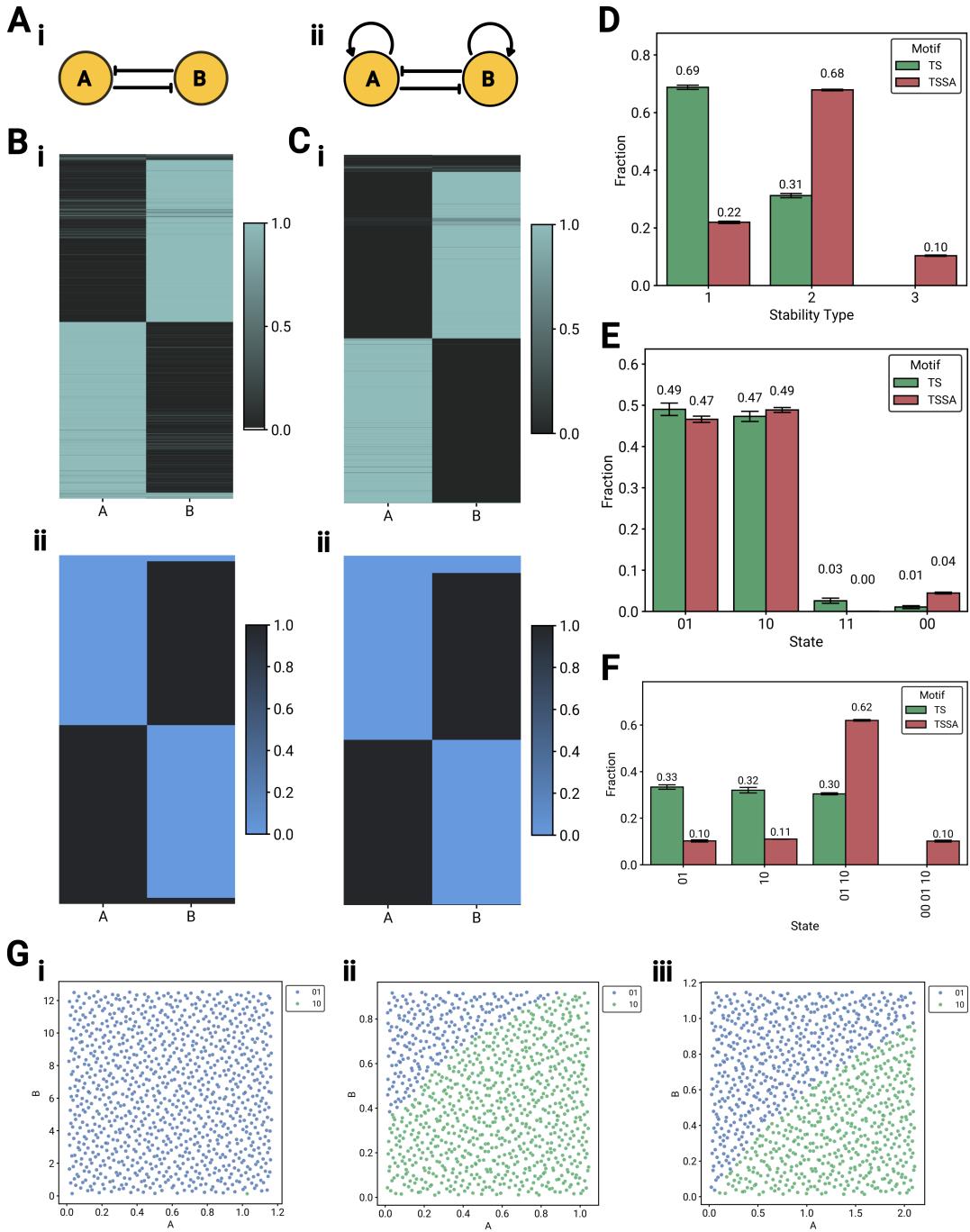


Figure 2: Comparison of toggle switch (TS) and toggle switch with self-activation (TSSA) motifs using the GRiNS framework. **A)** The toggle switch (TS) (i) and the toggle switch with self-activation (ii). **B)** Sample of expression profiles of TS nodes. i) G/k normalised expression ii) Discretised levels of the normalised expression values. **C)** Sample of expression profiles of TSSA nodes. i) G/k normalised expression ii) Discretised levels of the normalised expression values. **D)** Distribution of the parameters according to their multistability for TS and TSSA. **E)** Distributions of the steady states of TS and TSSA. **F)** Comparison of the multistable steady state frequencies between TS and TSSA. **G)** Initial condition (G/k normalised) mapping for three different bistable parameter sets.

conditions of the node values, colored by the corresponding steady-state they reach (Figure 2G, i-iii). This functionality—to track each initial condition and the steady state it leads to—can be combined with parameter sampling to investigate how the steady-state landscape changes as a specific parameter is varied. Since the initial conditions remain the same across parameter sweeps, the individual data points can be stacked, enabling more detailed analysis. In conclusion, the comparison between the TS and TSSA motifs revealed that the addition of self-activation on the nodes increased the ability to show bistability between the two mutually exclusive states of 01 and 10.

3.2 EMT Network - Comparison between RACIPE and Ising Boolean Formalisms

We now move on to a larger network—the EMT (epithelial-to-mesenchymal transition) network, which consists of 22 genes and 82 interactions (Figure 3A) [22]. This case study demonstrates how the same methods used for TS and TSSA can be scaled to larger networks and what additional insights can be gained from such simulations. The EMT network governs the transition of epithelial cells into mesenchymal cells, a process often dysregulated in cancer [30, 31, 32]. This dysregulation can result in hybrid phenotypes that are more challenging to treat and are associated with increased resistance to therapies [33]. We simulated this network in triplicate, sampling 10,000 parameter sets and 100 initial conditions for each replicate. One way to assess whether the number of parameters and initial conditions chosen is sufficient is to examine the distribution of coarse-grained steady states across replicates. If the standard deviations across replicates are low, it suggests that the sampling is consistent and that results are reliable, i.e., different replicates yield similar steady-state distributions. In our case, the low variability observed confirms that the selected number of parameter sets and initial conditions adequately capture the diversity of the GRN’s steady states (Figure 3B).

The EMT network contains several input nodes (with only outgoing edges) and output nodes (with only incoming edges). Due to how RACIPE is set up, randomly sampling initial conditions means that input nodes may be turned on or off arbitrarily without any regulation by other genes. Similarly, output nodes reflect the status of their upstream regulators. To make the analysis more focused and meaningful, we excluded the expression profiles of these input and output nodes from further analysis. An alternative way to address this issue is to modify the initial conditions of the input nodes based on the context of the simulation. For instance, one could set the input nodes connected to mesenchymal genes to high values, thereby simulating the influence of external factors on the network. In such scenarios, the limitations associated with randomly chosen initial conditions would no longer apply, and excluding these nodes from the analysis would not be necessary.

One notable observation from the coarse-grained steady-state distributions is that epithelial genes tend to be strongly co-expressed, whereas mesenchymal genes do not show consistent co-expression (Figure 3B). Self-inhibitory interactions in the network, like the ones present on SNAI1 and ZEB1, may partly explain this variability among mesenchymal genes. Despite the EMT network’s size, most parameter sets produced majorly bistable followed by monostable outcomes, suggesting strong coordination among genes leading to a constrained state space (Figure 3C), a pattern consistent with previous studies [34, 15]. To explore these states further, we performed Principal Component Analysis (PCA) on the steady-state gene expression profiles normalized by their G/k ratios. As the coarse-

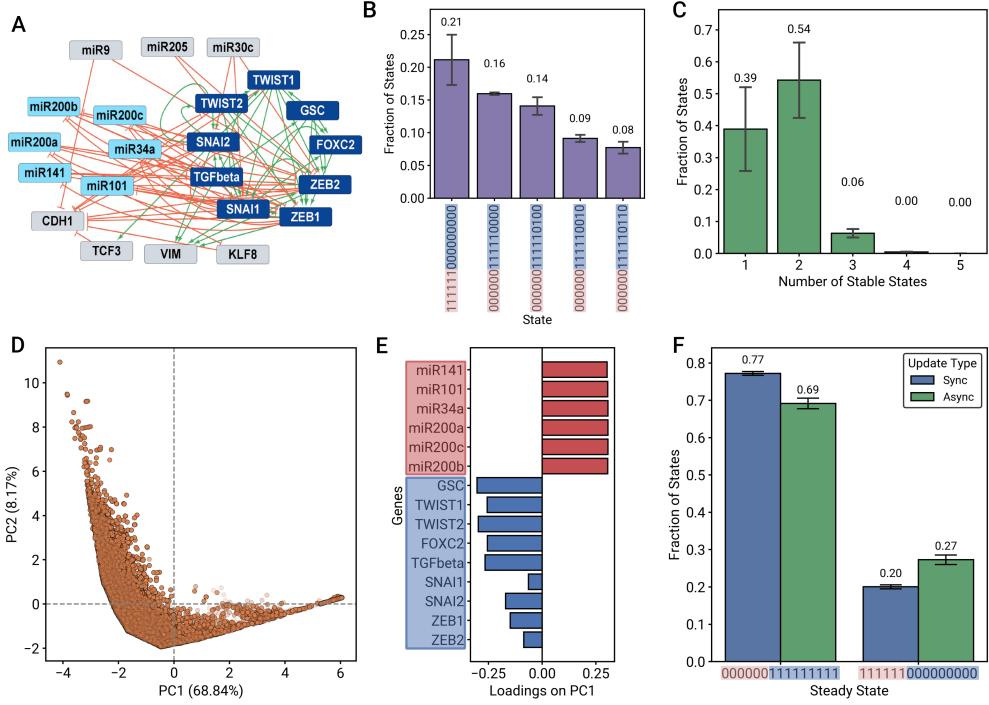


Figure 3: Analysis of the EMT gene regulatory network using RACIPE and Ising Boolean formalisms. **A)** Network diagram of the 22-node EMT GRN with 82 regulatory edges. Input and output nodes (marked grey) and excluded from analysis. **B)** Discretised steady-state distributions from RACIPE simulations. **C)** Distribution of multistability types from RACIPE simulations. **D)** PCA of G/k normalized steady states. **E)** PC1 loadings from the PCA. **F)** Steady state frequency distribution from synchronous and asynchronous Ising Boolean simulations.

grained analysis indicated, PCA results confirm that epithelial genes cluster tightly as seen by the distribution of the points on the positive side of the PC1 axis, reflecting strong co-expression (Fig. 3D). This trend is also confirmed by the PCA loadings of the first principal component: epithelial genes have similar values, while mesenchymal genes exhibit more spread in their loadings (Fig. 3E). Overall, this analysis suggests that even though coarse-graining reduces the granularity of continuous data, it still preserves essential features and makes the interpretation of large network simulations more intuitive and tractable.

An alternative approach for handling such large networks is to use the Ising Boolean formalism. Due to its non-parametric nature and implementation via matrix multiplications, this method is faster and can quickly approximate the steady-state landscape of a network. This becomes especially important for networks like EMT or even larger ones, where the computational cost and VRAM limitations make solving large systems of ODEs impractical. We simulated the network using both synchronous and asynchronous Ising update modes. Both methods produced the same dominant states: one where epithelial genes are active and mesenchymal genes are inactive, and the other with the reverse pattern (Figure 3F). However, unlike RACIPE simulations, the Ising formalism failed to capture intermediate or hybrid states. This limitation stems from the structure of the formalism, since no parameters are assigned to nodes or edges, the nuances that give rise to hybrid states are not preserved. Regarding the differences between the two update

rules, synchronous updates produce deterministic outcomes and do not account for multistability. On the other hand, asynchronous updates can lead to different outcomes from the same initial conditions due to the randomness in update sequences. Consequently, the variation in state frequencies between the two modes results in the asynchronous mode showing a greater number of unique states as compared to the synchronous mode, this also results in a lower value of the fraction of the most frequent state seen as the state distribution is relatively flatter, especially for the larger networks.

When comparing the results from RACIPE and the Ising formalism, both approaches consistently showed a higher frequency of the epithelial-low/mesenchymal-high state (although the states where mesenchymal genes are on do not tend to have all of them being on together) than the opposite of epithelial-high/mesenchymal-low states (Figure 3B; Figure 3F). Overall, the ODE-based approach offers richer data and enables more detailed analyses than the Ising formalism. However, simpler models such as the Ising Boolean formalism for large-scale exploratory studies can still offer valuable approximations of a network’s state space composition, especially when many networks need to be analyzed and compared. The Ising formalism can be a good approximation of a network’s dynamics and act as an initial step toward understanding the network during analysis.

4 Conclusion

GRiNS uses parameter-agnostic simulation frameworks and sampling strategies to model GRNs and capture their dynamics under various conditions. Implementing these methods in an accessible language like Python coupled with modular functions allows users to customize their simulations and explore the dynamic properties of GRNs in a standardized and reproducible manner. Thanks to its GPU-based simulation, GRiNS can be easily integrated into machine learning pipelines for cases like network inference, where unbiased simulation methods are important to determine and evaluate the predicted networks. The documentation and installation instructions for GRiNS can be found at <https://moltenecdysone09.github.io/GRiNS/>.

5 Availability and Requirements

- **Project name:** Gene Regulatory Interaction Network Simulator (GRiNS)
- **Project home page:** <https://moltenecdysone09.github.io/GRiNS/>
- **Operating system(s):** Platform Independent
- **Programming language:** *Python3*
- **License:** GNU GPL-3.0 License

Competing Interests

The authors declare no competing interests.

Author Contributions Statement

M.K.J. and P.H. conceived the idea of the simulation package. P.H. and H.B.V. wrote and developed the package. M.K.J. and P.H. wrote and reviewed the manuscript

Acknowledgments

M.K.J. received support from Param Hansa Philanthropies. P.H. and H.B.V. were supported by their respective Prime Minister's Research Fellowships (PMRF) awarded by the Government of India.

References

- [1] Qiuyue Yuan and Zhana Duren. Inferring gene regulatory networks from single-cell multiome data using atlas-scale external data. *Nature Biotechnology*, 43(2):247–257, February 2025.
- [2] Masato Ishikawa, Seiichi Sugino, Yoshie Masuda, Yusuke Tarumoto, Yusuke Seto, Nobuko Taniyama, Fumi Wagai, Yuhei Yamauchi, Yasuhiro Kojima, Hisanori Kiryu, Kosuke Yusa, Mototsugu Eiraku, and Atsushi Mochizuki. RENGE infers gene regulatory networks using time-series single-cell RNA-seq data with CRISPR perturbations. *Communications Biology*, 6(1):1–14, December 2023.
- [3] Pau Badia-i-Mompel, Lorna Wessels, Sophia Müller-Dott, Rémi Trimbour, Riccardo O. Ramirez Flores, Ricard Argelaguet, and Julio Saez-Rodriguez. Gene regulatory network inference in the era of single-cell multi-omics. *Nature Reviews Genetics*, 24(11):739–754, November 2023.
- [4] Roberto Barbuti, Roberta Gori, Paolo Milazzo, and Lucia Nasti. A survey of gene regulatory networks modelling methods: From differential equations, to Boolean and qualitative bioinspired models. *Journal of Membrane Computing*, 2(3):207–226, October 2020.
- [5] Mariana Gómez-Schiavon, Isabel Montejano-Montelongo, F. Sophia Orozco-Ruiz, and Cristina Sotomayor-Vivas. The art of modeling gene regulatory circuits. *npj Systems Biology and Applications*, 10(1):1–6, May 2024.
- [6] Tomáš Gedeon. Network topology and interaction logic determine states it supports. *npj Systems Biology and Applications*, 10(1):1–10, August 2024.
- [7] Guy Karlebach and Ron Shamir. Modelling and analysis of gene regulatory networks. *Nature Reviews Molecular Cell Biology*, 9(10):770–780, October 2008.
- [8] Federico Bocci, Dongya Jia, Qing Nie, Mohit Kumar Jolly, and José Onuchic. Theoretical and computational tools to model multistable gene regulatory networks. *Reports on Progress in Physics*, 86(10):106601, August 2023.
- [9] Mohit Kumar Jolly and Herbert Levine. Computational systems biology of epithelial-hybrid-mesenchymal transitions. *Current Opinion in Systems Biology*, 3:1–6, June 2017.

- [10] Weixu Wang, Zhiyuan Hu, Philipp Weiler, Sarah Mayes, Marius Lange, Jingye Wang, Zhengyuan Xue, Tatjana Sauka-Spengler, and Fabian J. Theis. RegVelo: Gene-regulatory-informed dynamics of single cells. *bioRxiv*, 2024.
- [11] Aditya Pratapa, Amogh P. Jalihal, Jeffrey N. Law, Aditya Bharadwaj, and T. M. Murali. Benchmarking algorithms for gene regulatory network inference from single-cell transcriptomic data. *Nature Methods*, 17(2):147–154, February 2020.
- [12] Malvina Marku and Vera Pancaldi. From time-series transcriptomics to gene regulatory networks: A review on inference methods. *PLOS Computational Biology*, 19(8):e1011254, August 2023.
- [13] Philipp Städter, Yannik Schälte, Leonard Schmiester, Jan Hasenauer, and Paul L. Stapor. Benchmarking of numerical integration methods for ODE models of biological systems. *Scientific Reports*, 11(1):2696, January 2021.
- [14] Fabian Fröhlich and Peter K. Sorger. Fides: Reliable trust-region optimization for parameter estimation of ordinary differential equation models. *PLOS Computational Biology*, 18(7):e1010322, July 2022.
- [15] Bin Huang, Mingyang Lu, Dongya Jia, Eshel Ben-Jacob, Herbert Levine, and Jose N. Onuchic. Interrogating the topological robustness of gene regulatory circuits by randomization. *PLOS Computational Biology*, 13(3):e1005456, March 2017.
- [16] Bin Huang, Dongya Jia, Jingchen Feng, Herbert Levine, José N. Onuchic, and Mingyang Lu. RACIPE: A computational tool for modeling gene regulatory circuits using randomization. *BMC Systems Biology*, 12(1):74, June 2018.
- [17] Anupam Dey and Debashis Barik. Potential Landscapes, Bifurcations, and Robustness of Tristable Networks. *ACS Synthetic Biology*, 10(2):391–401, February 2021.
- [18] Xinyu He, Ruoyu Tang, Jie Lou, and Ruiqi Wang. Identifying key factors in cell fate decisions by machine learning interpretable strategies. *Journal of Biological Physics*, 49(4):443–462, December 2023.
- [19] Fangting Li, Tao Long, Ying Lu, Qi Ouyang, and Chao Tang. The yeast cell-cycle network is robustly designed. *Proceedings of the National Academy of Sciences*, 101(14):4781–4786, April 2004.
- [20] Roberto Barbuti, Roberta Gori, and Paolo Milazzo. Encoding Boolean networks into reaction systems for investigating causal dependencies in gene regulation. *Theoretical Computer Science*, 881:3–24, August 2021.
- [21] Lingyu Li, Liangjie Sun, Guangyi Chen, Chi-Wing Wong, Wai-Ki Ching, and Zhi-Ping Liu. LogBTF: Gene regulatory network inference using Boolean threshold network model from single-cell gene expression data. *Bioinformatics*, 39(5):btad256, May 2023.
- [22] Francesc Font-Clos, Stefano Zapperi, and Caterina A. M. La Porta. Topography of epithelial–mesenchymal plasticity. *Proceedings of the National Academy of Sciences*, 115(23):5902–5907, June 2018.

- [23] James Bradbury, Roy Frostig, Peter Hawkins, Matthew James Johnson, Chris Leary, Dougal Maclaurin, George Necula, Adam Paszke, Jake VanderPlas, Skye Wanderman-Milne, and Qiao Zhang. JAX: Composable transformations of Python+NumPy programs, 2018.
- [24] Patrick Kidger. On Neural Differential Equations. *arXiv*, 10.48550/arXiv.2202.02435, February 2022.
- [25] Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, Stéfan J. van der Walt, Matthew Brett, Joshua Wilson, K. Jarrod Millman, Nikolay Mayorov, Andrew R. J. Nelson, Eric Jones, Robert Kern, Eric Larson, C. J. Carey, İlhan Polat, Yu Feng, Eric W. Moore, Jake VanderPlas, Denis Laxalde, Josef Perktold, Robert Cimrman, Ian Henriksen, E. A. Quintero, Charles R. Harris, Anne M. Archibald, Antônio H. Ribeiro, Fabian Pedregosa, and Paul van Mulbregt. SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nature Methods*, 17(3):261–272, March 2020.
- [26] Merja Heinäniemi, Matti Nykter, Roger Kramer, Anke Wienecke-Baldacchino, Lasse Sinkkonen, Joseph Xu Zhou, Richard Kreisberg, Stuart A. Kauffman, Sui Huang, and Ilya Shmulevich. Gene-pair expression signatures reveal lineage control. *Nature Methods*, 10(6):577–583, June 2013.
- [27] Sui Huang, Yan-Ping Guo, Gillian May, and Tariq Enver. Bifurcation dynamics in lineage-commitment in bipotent progenitor cells. *Developmental Biology*, 305(2):695–713, May 2007.
- [28] Mitra Mojtabaei, Alexander Skupin, Joseph Zhou, Ivan G. Castaño, Rebecca Y. Y. Leong-Quong, Hannah Chang, Kalliopi Trachana, Alessandro Giuliani, and Sui Huang. Cell Fate Decision as High-Dimensional Critical State Transition. *PLOS Biology*, 14(12):e2000640, December 2016.
- [29] Raúl Guantes and Juan F. Poyatos. Multistable Decision Switches for Flexible Control of Epigenetic Differentiation. *PLOS Computational Biology*, 4(11):e1000235, November 2008.
- [30] Pasquale Simeone, Marco Trerotola, Julien Franck, Tristan Cardon, Marco Marchisio, Isabelle Fournier, Michel Salzet, Michele Maffia, and Daniele Vergara. The multiverse nature of epithelial to mesenchymal transition. *Seminars in Cancer Biology*, 58:1–10, October 2019.
- [31] Anushka Dongre and Robert A. Weinberg. New insights into the mechanisms of epithelial–mesenchymal transition and implications for cancer. *Nature Reviews Molecular Cell Biology*, 20(2):69–84, February 2019.
- [32] Alexandre Francou and Kathryn V. Anderson. The Epithelial-to-Mesenchymal Transition in Development and Cancer. *Annual Review of Cancer Biology*, 4(Volume 4, 2020):197–220, March 2020.
- [33] Mohit Kumar Jolly, Sendurai A. Mani, and Herbert Levine. Hybrid epithelial/mesenchymal phenotype(s): The ‘fittest’ for metastasis? *Biochimica et Biophysica Acta (BBA) - Reviews on Cancer*, 1870(2):151–157, December 2018.

- [34] Kishore Hari, Pradyumna Harlapur, Aashna Saxena, Kushal Haldar, Aishwarya Girish, Tanisha Malpani, Herbert Levine, and Mohit Kumar Jolly. Low dimensionality of phenotypic space as an emergent property of coordinated teams in biological regulatory networks. *iScience*, 28(2):111730, February 2025.