Computational techniques for analysis of genetic network dynamics

Calin Belta [a], Joel M. Esposito [b],
Jongwoo Kim [c], and Vijay Kumar [c]
[a] Drexel University, Philadelphia, PA
[b] US Naval Academy, Annapolis, MD
[c] University of Pennsylvania, Philadelphia, PA
calin@drexel.edu, esposito@usna.edu,
jwk,kumar@grasp.cis.upenn.edu

November 11, 2003

Abstract

In this paper we propose modeling and analysis techniques for genetic networks that provide biologists with an insight into the dynamics of the system. These techniques are scalable to high dimensions and can incorporate uncertainty (partial knowledge of kinetic parameters and state uncertainty). We are not only interested in developing mathematical models and simulating them, but in proving biologically significant properties. Central to our modeling approach is the framework of hybrid systems and our analysis tools derived from formal analysis of such systems. Given a set of states characterizing a property of interest $\mathcal{P}$, we present the Multi-Affine Rectangular Partition (MARP) algorithm for the construction of a set of infeasible states $\mathcal{I}$ that will never reach $\mathcal{P}$ and the Rapidly Exploring Random Forest of Trees (RRFT) algorithm for the construction of a set of feasible states $\mathcal{F}$ that will reach $\mathcal{P}$. We apply these methods to understand the genetic interactions involved in the phenomenon of luminescence production in the marine bacterium *V. fischeri*.

1 Introduction

The recent completion of a draft of the human genome and the sequencing of several other organisms provides a vast amount of genomic data for advancing our understanding of fundamental biological processes. However, in order to understand different cellular behaviors such as differentiation, response to environmental signals, and cell-to-cell communication, we need to study the regulatory systems determining expressions of genes. This is usually a complex process, which can be regulated at several stages such as transcription (the best

studied form of regulation), translation, and post translational modification of proteins. An example of transcriptional regulation is repression: a regulatory molecule binds to a regulatory site of some gene preventing the RNA polymerase from transcribing the gene. The number of regulating factors is usually large, and it involves proteins (products of other genes and possibly of the gene itself), RNA, and other molecules. A collection of interacting genes, their products, and some other molecules involved in the regulation of the genes form a *genetic regulatory network*.

The traditional approach to modeling of genetic networks leads to highly nonlinear systems of differential equations for which analytical solutions are not normally possible. The only alternative for analysis is numerical simulation. One way to work around the difficulties of the nonlinearities is to use simplified, approximate models. Existing work focuses on very low-dimensional genetic networks. Decoupled piecewise linear differential equations (PLDE) are considered in [22, 40], where gene regulation is modeled as a discontinuous step function and chemical reactions are ignored. This (over)simplified approach to modeling allows for interesting qualitative analysis [15]. An even more radical idealization is obtained if the state of a gene is abstracted to a Boolean variable and the interaction among elements to Boolean functions, as in Boolean networks [28]. Other types of simplified approaches combining logical and continuous aspects include generalized logical formalisms [50] and qualitative differential equations [34]. The highest level of abstraction is achieved in the knowledge-based, or rule-based formalism [10].

While amenable for interesting analysis, the methods mentioned above are based on assumptions which disregard important biochemical phenomena. Most of them only capture protein dynamics but cannot accommodate chemical reactions [28]. When the regulatory systems are not spatially homogeneous, partial differential equations and other spatially distributed models such as reaction-diffusion equations [29] and the gene circuit method [43] can be used. The modeling approach based on differential equations implicitly assumes that the concentrations of the species in the network vary continuously and deterministically. The continuity assumption is compromised in cases when the number of molecules of a certain specie is small or due to fluctuations in the timing of cellular events. The use of discrete stochastic models is proposed in [21, 39, 13]. However, these methods are computationally expensive and cannot handle high-dimensional systems.

Our modelling approach is deterministic and based on *hybrid systems* [38], *i.e.,* systems in which discrete events are combined with continuous differential equations to capture the switching behavior that is observed in phenomena such as transcription, protein-protein interactions, and cell division and growth. We also propose the use of hybrid system as the natural framework giving a global description of a biological system described locally around operating points by simpler dynamics, which are easier to approach for analysis. Our own work using hybrid systems to model, simulate and perform preliminary analysis on low-dimensional genetic networks is given [3, 7, 8, 4, 2].

We are interested in developing general modeling, simulation, and analysis

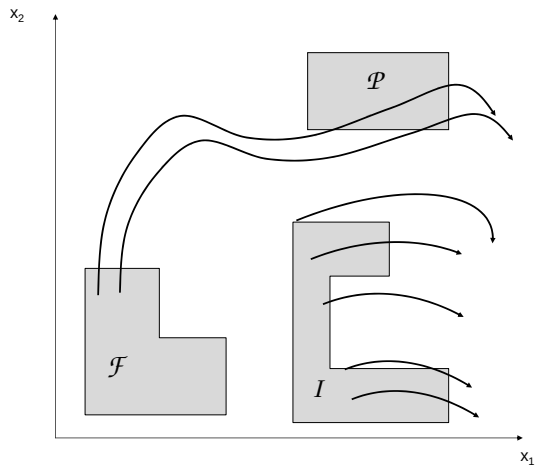

Figure 1: The three sets of interest: the property set $\mathcal{P}$, the infeasible set $\mathcal{I}$ from which no trajectory can enter $\mathcal{P}$, and the feasible set $\mathcal{F}$ from which there exists at least one trajectory which can enter $\mathcal{P}$.

techniques for metabolic and genetic networks to deal with partial knowledge of kinetic information, and possibly innacurate experimental data. Our ultimate goal is to create tools enabling us to answer biologically significant questions of the type: "if an organism is initially in a state partially or completely described by certain ranges of metabolite and enzyme concentrations, kinetic parameters, and levels of activation of genes, describe the set of states that the organism can reach while time evolves". Or, "describe the states with the property that if the system starts in any of them it will never reach an undesired state as time flows". The undesired state could correspond, for example, to a certain disease.

We denote the set associated with *properties* of interest as $\mathcal{P}$. To answer the questions formulated above, we are interested in determining two disjoint types of sets of initial conditions that can be associated with the set $\mathcal{P}$. See Figure 1. First, we will be interested in characterizing *feasible sets* $\mathcal{F}$, consisting of initial conditions that make it possible for the system to enter the set $\mathcal{P}$ under parameters uncertainties and noise in the model. We are also interested in the *infeasible sets* $\mathcal{I}$, from which it is impossible to enter the set $\mathcal{P}$. Knowledge of $\mathcal{F}$ may assist in experiment design; while a knowledge of the set $\mathcal{I}$, is particularly useful for model validation. If experimental data indicate the system enters $\mathcal{P}$ and if experimental conditions were known to lie in set $\mathcal{I}$, one could prove the model to be inconsistent with the experimental data. In many ways the approaches to generating the two sets, and the information they encode, are complementary – neither approach alone is complete. Together however, knowledge of the sets $\mathcal{F}$ and $\mathcal{I}$ can enable us to make some powerful assertions about the system's behavior.

The connection between biomolecular networks and robotic systems exists on two levels. From a modelling point of view robotic systems share many of the

salient features of biological system models described above. Just as the rate equations for biomolecular are known to qualitatively switch based on the presence or absence of various inhibitor genes, robotic systems often employ different controllers and estimators in different regimes [14] and their dynamics switch based on the contact mechanics of rolling and sliding. Therefore, both types systems can be modelled as *hybrid* systems. Indeed there are several papers that use hybrid models for molecular and cellular networks [3, 7, 8, 4, 2, 22, 15]. In addition, many robotic systems consist of multi-agent teams, and therefore the interactions and messaging among the team members must be taken into account and many of the system properties are distributed spatially. Multi-cell networks behave in much the same way. Finally the significant modeling uncertainty, which is central to our discussion and analysis of biological systems, is a common theme in mobile robotics operating in unstructured environments.

Perhaps a deeper connection between the two fields exists at the level of the types of problems we seek to solve. The problem of finding sets of states $\mathcal{F}$, from which the system may reach $\mathcal{P}$, is similar to the *motion planning problem* in robotics where the goal is, given a robot with dynamics and constraints (obstacles), to find a path or trajectory (if one exists) from the starting configuration to the goal configuration. Determining an infeasible set $\mathcal{I}$, from which it is impossible to reach the property set $\mathcal{P}$ is closely related to trajectory generation, controllability and steering. As one can imagine, the literature on simulation and verification of hybrid systems is also particularly relevant to our discussion [46, 25, 35, 49, 5, 1, 11, 42, 41, 12].

In this paper we develop methodologies for finding the sets shown in Figure 1. To construct infeasible sets $\mathcal{I}$, we propose the Multi-Affine Rectangular Partitioning (MARP) algorithm, which exploits the specific properties of hybrid models of genetic networks: they have rectangular invariants because different behaviors are specified as functions of ranges of concentrations of regulatory species and the vector fields have product type non-linearities due to chemical reactions. For these types of systems, which we denote as *rectangular multi-affine hybrid systems*, given a set $\mathcal{P}$, we develop a computationally efficient method to construct an infeasible set $\mathcal{I}$ based on evaluating the vector field at the vertices of the rectangular invariants. Given a specific initial condition, the popular Rapidly Exploring Random Tree (RRT) algorithm from the motion planning literature provides a natural way for one to determine if a set should be included in $\mathcal{F}$, by searching for a trajectory which reaches $\mathcal{P}$, over all possible noise functions (i.e., inputs). However, it does not provide a natural way to address time invariant uncertainty such as unknown initial conditions or rate constants. We introduce an algorithm called the Rapidly Exploring Random Forest of Trees (RRFT) which samples the set of possible initial conditions using a Monte Carlo type approach and "plants" a RRT at each sample point. Each RRT can be grown on a different processor. We then introduce various measures of growth and coverage for RRT's which can be used to allocate computational resources effectively among the set of trees. If the growth of a given RRT has slowed below a certain threshold, it can be terminated and another RRT is planted dynamically. The process continues until a trajectory which reaches $\mathcal{P}$

is discovered or until the sample set is sufficiently dense that one can conclude the existence of such a trajectory is unlikely. The RRFT algorithm allows one to address the reachability problem probabilistically for complex high-dimensional systems having both time-varying and time-invariant uncertainty.

The rest of the paper is organized as follows. In Section 2, we introduce the hybrid system modelling paradigm and provide the basic definitions that will be used throughout the paper. In Section 3, we exploit the particular structure of the hybrid models of genetic networks to derive a computationally attractive algorithm (Multi-Affine Rectangular Partitioning - MARP) to compute *infeasible sets* $\mathcal{I}$. We then describe a randomized algorithm (Rapidly Exploring Random Forest of Trees - RRFT) that is used to intelligently explore a large state space to identify points in the *feasible sets* $\mathcal{F}$ in Section 4. The usefulness of the two algorithms is illustrated in Section 5, where we study the phenomenon of bioluminescence production in the marine bacterium *V. fischeri* by analyzing the corresponding genetic network. The paper concludes with final remarks and directions of future research in Section 6.

2 Hybrid system modelling of genetic networks

Hybrid systems are dynamical systems with both discrete and continuous state changes [38]. We are interested in a special class of hybrid system, called *Switched systems*, which are defined as having different dynamics in different nonoverlapping regions of the state space. In our view, hybrid and switched systems are appropriate and attractive for modeling the dynamics of biomolecular networks for two main reasons:

Hybrid systems are global descriptions from simpler local models Computationally attractive formalisms for modelling biomolecular networks such as linearizations, half-systems [48], synergistic (S) systems [47], generalized mass - action (GMA) [45]), and power law [24] are only valid locally around operating points. For example, the S-systems can be thought of linearizations of "real" systems in logarithmic coordinates [47]. Then, in this case, a global description of the network is a collection of regions with different posynomial vector fields in each region, therefore a hybrid system. The specific nonlinearites of dynamics in each region are simpler than the dynamics of a global continuous description, and easier to approach for analysis.

Hybrid systems capture discrete events Discrete dynamics are necessary to capture switching behavior that is observed in phenomena such as transcription, protein-protein interactions, and cell division and growth. Consider, for example, the case when a metabolite from the network regulates the production of a metabolic enzyme expressed from a gene with a strong promoter. Then, the gene can be "on" and "off", which induces two different dynamics of the network, as a function of the concentration of the metabolite.

Formally, hybrid systems are defined as tuples

$$HS = (Q, X, \mathbf{X}^0, I, T, F) \tag{1}$$

where Q is a finite set of discrete variables, X is the set of continuous variables x, $\mathbf{X}$ is the set of all evaluations of x over the corresponding domains, $\mathbf{Q}$ is the countable set of discrete states, called modes, or locations, $\mathbf{X}^0 \subset \mathbf{Q} \times \mathbf{X}$ is a set of initial states, I is a map which assigns to each discrete location in $\mathbf{Q}$ an invariant set, $T \subset \mathbf{Q} \times \mathbf{X} \times \mathbf{Q}$ is a set of discrete transitions, and $F : \mathbf{Q} \to (\mathbf{X} \to T\mathbf{X})$ is a mapping that specifies the continuous (possibly time dependent) flow in each discrete state.

We focus on vector fields f which are products of the state components to capture the nonlinearities which are specific to dynamics of chemical reactions. To take into account possible modelling noise and to accommodate non-deterministic modelling approaches, we also allow for an additive noise term $\nu(t)$ in the vector field. Therefore, as suggested by [31], the vector field associated with the map F takes the form $F = f(x) + \nu(t)$. The form of f is made precise in the next subsection, while ν is discussed in Section 4.

We are also interested in constant parameters that might be unknown. Examples of these parameters are binding constants and other constants determining reaction rate kinetics. While these constants are known to lie within known bounds, their exact values are often unknown. It is useful to note that these unknown constants can be viewed as state variables with trivial dynamics, *e.g.*, $x = c,\ \dot{x} = 0$, while the corresponding projection of $\mathbf{X}^0$ characterizes the known bounds. Thus, our definition of HS in (1) allows for unknown parameters that lie in some specified set.

Finally, since molecular networks are qualitatively described in terms of ranges of concentrations of the involved species, a rectangular partition of the state space is naturally induced and the invariants are rectangular.

2.1 Rectangular multi-affine hybrid systems

Rectangular multi-affine hybrid systems are characterized by rectangular invariants and multi-affine continuous dynamics.

Definition 2.1 (Multi-affine function) *A multi-affine function $f : \mathbb{R}^N \longrightarrow \mathbb{R}^N$ is a polynomial in the indeterminates $x_1, \ldots, x_N$ with the property that the degree of f in any of the indeterminates $x_1, \ldots, x_N$ is less than or equal to 1. Stated differently, f has the form*

$$f(x_1, \ldots, x_N) = \sum_{i_1, \ldots, i_N \in \{0,1\}} c_{i_1, \ldots, i_N} x_1^{i_1} \cdots x_N^{i_N}, \tag{2}$$

with $c_{i_1, \ldots, i_N} \in \mathbb{R}^N$ for all $i_1, \ldots, i_N \in \{0, 1\}$ and using the convention that if $i_k = 0$, then $x_k^{i_k} = 1$.

An N-dimensional rectangle in $\mathbb{R}^N$ is characterized by two vectors $a = (a_1, \ldots, a_N) \in \mathbb{R}^N$ and $b = (b_1, \ldots, b_N) \in \mathbb{R}^N$, with the property that $a_i < b_i$ for $i = 1, \ldots, N$:

$$R_N(a, b) = \{x = (x_1, \ldots, x_N) \in \mathbb{R}^N \mid i = 1, \ldots, N : \ a_i \leq x_i \leq b_i\}. \quad (3)$$

x_i, $i = 1, \ldots, N$ are species concentrations and are restricted to the positive quadrant. Also, there are practical upper bounds on the concentration of each specie. Therefore, the set $\mathbf{X}$ as in (1) is usually specified as an N-rectangle. A rectangular partition of X is defined as follows. Each axis Ox_i, $i = 1, \ldots, N$ is divided into $n_i \geq 1$ intervals by the thresholds $0 = \theta_i^0 < \theta_i^1 < \ldots < \theta_i^{n_i}$. The j^{th} interval on the Ox_iaxis, $i = 1, \ldots, N$ is therefore defined as $\theta_i^{j-1} \leq x_i < \theta_i^j$, $j = 1, \ldots, n_i$. By convention, $\theta_i^0 = 0$ and $\theta_i^{n_i}$ is an upper bound giving a physical limit of x_i. The thresholds θ are defined as values of species concentrations for each the dynamics of the overall system changes. For example, they can be concentrations of regulatory species for which specific genes are turned 'on' and 'off'. The division of the axes determines a partition of the state space into $\prod_{i=1}^N n_i$ rectangles. If we let

$$a^{(q_1 \ldots q_N)} = (\theta_1^{q_1 - 1}, \ldots, \theta_N^{q_N - 1}) \in \mathbb{R}^N, \ b^{(q_1 \ldots q_N)} = (\theta_1^{q_1}, \ldots, \theta_N^{q_N}) \in \mathbb{R}^N, \quad (4)$$

for $q_i = 1, \ldots, n_i$, $i = 1, \ldots, N$, then an arbitrary rectangle in the partition is given by

$$R_N(a^{(q_1 \ldots q_N)}, b^{(q_1 \ldots q_N)}) = \{(x_1, \ldots, x_N) \in \mathbb{R}^N | \theta_i^{q_i - 1} \leq x_i \leq \theta_i^{q_i}, \ i = 1, \ldots, N\} \quad (5)$$

Due to the different levels of gene transcription activation and enzymatic action, in each of the rectangles the system evolves along specific multi-affine vector fields (2):

$$f^{(q_1 \ldots q_N)}(x_1, \ldots, x_N) = \sum_{i_1, \ldots, i_N \in \{0,1\}} c_{i_1, \ldots, i_N}^{(q_1 \ldots q_N)} x_1^{i_1} \cdots x_N^{i_N}, \quad (6)$$

where $x \in R_N(a^{(q_1 \ldots q_N)}, b^{(q_1 \ldots q_N)})$ and $c_{i_1, \ldots, i_N}^{(q_1 \ldots q_N)} \in \mathbb{R}^N$ for all $i_1, \ldots, i_N \in \{0, 1\}$ captures specific reaction rates.

Therefore, our models of biomolecular networks are hybrid systems (1) with the set of labels for the discrete states $Q = (q_1 \ldots q_N)$, the set of all $\prod_{i=1}^N n_i$ modes $\mathbf{Q} = \{(q_1 \ldots q_N)|, \ q_i = 1, \ldots, n_i, \ i = 1, \ldots, N\}$, X is the set of species symbols $x_1, \ldots, x_N$, the invariant $I(q_1 \ldots q_N)$ is the corresponding rectangle (5), and the map F has a deterministic part described by (6). A transition $((q_1 \ldots q_N), x, (q_1' \ldots q_N'))$ corresponds to the crossing of the boundary between rectangles $I(q_1 \ldots q_N)$ and $I(q_1' \ldots q_N')$ at state x.

Remark 2.2 *Due to the particular shape of the invariants, a convenient way of representing a rectangular multi-affine hybrid system (6), (5) is as a* simple graph *with $\prod_{i=1}^N n_i$ nodes. Node $(q_1 \ldots q_N)$ corresponds to rectangle $I(q_1 \ldots q_N) = R_N(a^{(q_1 \ldots q_N)}, b^{(q_1 \ldots q_N)})$ and has associated dynamics (6). An edge in the graph*

connects nodes corresponding to adjacent rectangles, i.e., *there is an edge between any pair of nodes that differ by a Hamming distance of 1. See Figure 7 for an example with* $N = 3$, $n_1 = n_2 = n_3 = 3$.

2.2 Set definitions

As stated before, we are interested in characterizing the properties of the system HS related to whether it can reach a given set of interest $\mathcal{P}$ described by a rectangle $I(p_1 \dots p_N)$, $(p_1 \dots p_N) \in \mathbf{Q}$.

Definition 2.3 (Infeasible set) *An infeasible set $\mathcal{I}$ is a collection of rectangles $I(q_1 \dots q_N)$, $(q_1 \dots q_N) \in \mathbf{Q}$ with the property that the system can never reach $I(p_1 \dots p_N)$ if it starts in any of the initial states contained in any of rectangles in $\mathcal{I}$.*

Definition 2.4 (Feasible set) *A feasible set $\mathcal{F}$ is a collection of rectangles $I(q_1 \dots q_N)$, $(q_1 \dots q_N) \in \mathbf{Q}$ with the property that they contain initial states that will reach $I(p_1 \dots p_N)$ in a pre-specified finite time interval.*

3 Computing an infeasible set $\mathcal{I}$

This section introduces the Multi-Affine System Partitioning (MARP) algorithm which uses the properties of hybrid multiaffine rectangular systems to construct infeasible sets $\mathcal{I}$, and extends some results presented in [7]. In the form presented in this paper, the algorithm can only be applied to deterministic vector fields where $F = f$. However, as explained in Section 1, it can easily accommodate rectangular parametric uncertainties: set-valued uncertainty in a constant parameter can be included in the set of initial conditions, $\mathbf{X}^0$.

3.1 Preliminaries

The MARP algorithm is based on the fact that the value of a multi-affine function (6) is uniquely determined everywhere in the rectangular invariant $I(q_1 \dots q_N)$ by its values at the vertices. Moreover, it is a convex combination of these values.

Formally, for an arbitrary rectangle (3), let

$$V_N(a, b) = \prod_{i=1}^{N} \{a_i, b_i\} \tag{7}$$

denote the set of its 2^N vertices. Let $\xi : \{a_1, \dots, a_N, b_1, \dots, b_N\} \longrightarrow \{0, 1\}$ be defined by

$$\xi(a_k) = 0, \quad \xi(b_k) = 1, \quad k = 1, \dots, N \tag{8}$$

Proposition 3.1 *A multi-affine function $f : R_N(a,b) \longrightarrow \mathbb{R}^N$ is a convex combination of its values $f(v_1,\ldots,v_N)$ at the vertices $V_N(a,b)$. Explicitly,*

$$f(x_1,\ldots,x_N) = \sum_{(v_1,\ldots,v_N)\in V_N(a,b)} \prod_{k=1}^{N} \left(\tfrac{x_k-a_k}{b_k-a_k}\right)^{\xi(v_k)} \left(\tfrac{b_k-x_k}{b_k-a_k}\right)^{1-\xi(v_k)} f(v_1,\ldots,v_N), \tag{9}$$

with

$$1 = \sum_{(v_1,\ldots,v_N)\in V_N(a,b)} \prod_{k=1}^{N} \left(\frac{x_k-a_k}{b_k-a_k}\right)^{\xi(v_k)} \left(\frac{b_k-x_k}{b_k-a_k}\right)^{1-\xi(v_k)} \tag{10}$$

where $(v_1,\ldots,v_N) \in V_N(a,b)$.

The proof of this proposition can be found in [7]. Since the projection of a multi-affine vector field along a given direction is a multi-affine function, an immediate consequence of Proposition 3.1 can be used to develop a computationally efficient algorithm for constructing infeasible sets for rectangular multi-affine hybrid systems:

Corollary 3.2 *The projection of a multi-affine vector field defined on a rectangle along a given direction is positive (negative) everywhere in the rectangle if and only if its projection along that direction is positive (negative) at the vertices.*

3.2 MARP algorithm

We assume that the piecewise defined vector field (6) (possibly non-differentiable), is continuous everywhere, *i.e.,* the vector fields in adjacent rectangles coincide on the common facet. A simple consequence of Corollary 3.2 can be used to qualitatively analyze the system.

Corollary 3.2 is applied to the facets of the N - rectangles (5) and to the projections of the vector fields along the corresponding outer normals. Each facet is a $N-1$-rectangle. An infeasible set $\mathcal{I}$ can be built by defining an orientation for the simple graph of the network defined in Remark 2.2. We allow for both unidirectional and bidirectional edges in the oriented graph. The semantics of the orientation are defined as follows. Let $(q_1 \ldots q_N)$ and $(q'_1 \ldots q'_N)$ be two adjacent nodes in the graph and $I(q_1 \ldots q_N)$ and $I(q'_1 \ldots q'_N)$ the corresponding adjacent rectangles. A unidirectional edge from $(q_1 \ldots q_N)$ to $(q'_1 \ldots q'_N)$ means that there exists at least one trajectory originating in $I(q_1 \ldots q_N)$ that enters into $I(q'_1 \ldots q'_N)$ through the separating facet, and there is no trajectory starting in $I(q'_1 \ldots q'_N)$ going to $I(q_1 \ldots q_N)$ through that facet. A bidirectional edge insures the existence of at least one trajectory originating in $I(q_1 \ldots q_N)$ eneteering in $I(q'_1 \ldots q'_N)$ and at least one trajectory originating in $I(q'_1 \ldots q'_N)$ entering into $I(q_1 \ldots q_N)$.

Note that, in the oversimplified description above, Algorithm 1 seems inefficient. Indeed, if we apply it to all the rectangles in the partition, most of

Algorithm 1 Define an oriented graph

for each node $(q_1 \dots q_N)$, $q_i = 1, \dots, n_i$, $i = 1, \dots, N$ **do**
  for each incident edge **do**
    for each vertex of the corresponding facet **do**
      calculate the projection of $f^{(q_1 \cdots q_N)}$ along the outer normal of the facet
    end for
    if the projections are positive at all vertices **then**
      the edge is unidirectional oriented out of $(q_1 \dots q_N)$
    end if
    if the projections are negative at all vertices **then**
      the edge is unidirectional oriented towards $(q_1 \dots q_N)$
    end if
    if there is a sign change among the projections at the vertices **then**
      the edge is bidirectional
    end if
  end for
end for

the vertices are visited more than ones, and this is not necessary because the vector fields in adjacent rectangles match on the separating facet. A more efficient description would require more complicated notation and a more detailed discussion which we omit because it is peripheral to the main ideas in the paper.

Using the oriented graph, we can now construct an infeasible set $\mathcal{I}$. Let $\mathcal{P} = I(p_1 \dots p_N)$ denote the target rectangle, or, equivalently, $(p_1 \dots p_N)$ is the target node in the graph. The following algorithm constructs a set $\mathcal{R}$ of nodes with the property that if the system starts in any of the corresponding rectangles, then it *may* be possible to reach $\mathcal{P}$. The complement of this set is an infeasible set $\mathcal{I}$.

Algorithm 2 Construct an infeasible set $\mathcal{I}$

initialize $\mathcal{R}$ with $\mathcal{P}$
repeat
  for each element $(q_1 \dots q_N)$ of $\mathcal{R}$ **do**
    for all incident nodes $(q'_1 \dots q'_N)$ connected with an edge (uni or bi-directional) oriented towards $(q_1 \dots q_N)$ **do**
      if $(q'_1 \dots q'_N)$ is not already in $\mathcal{R}$ **then**
        add $(q'_1 \dots q'_N)$ to $\mathcal{R}$
      end if
    end for
  end for
until cardinality of $\mathcal{R}$ increases
$\mathcal{I}$:= complement of $\mathcal{R}$ with respect to the set $\mathbf{Q}$ of all nodes

Remark 3.3 *Algorithm 2 for the construction of the infeasible set $\mathcal{I}$ might be too conservative,* i.e., *the set $\mathcal{I}$ might be unnecessarily small. Indeed, our method guarantees the existence of a trajectory from a rectangle $I(q_1 \dots q_N)$ to an adjacent rectangles $I(q'_1 \dots q'_N)$ if the (unidirectional or bidirectional) edge between the corresponding nodes in the oriented graph has an arrow from $(q_1 \dots q_N)$ to $(q'_1 \dots q'_N)$. But if there is an edge from $(q_1 \dots q_N)$ to $(q'_1 \dots q'_N)$ and also an edge from $(q'_1 \dots q'_N)$ to $(q''_1 \dots q''_N)$, we cannot guarantee that there is a trajectory of the system from $I(q_1 \dots q_N)$ to $I(q''_1 \dots q''_N)$. In our analysis, we simply say that there might be a trajectory from $I(q_1 \dots q_N)$ to $I(q''_1 \dots q''_N)$.*

Remark 3.4 *The above algorithm can be easily extended to construct infeasible sets $\mathcal{I}$ under rectangular parameter uncertainties. This is possible because the parameters c capturing kinetic constants enter the vector fields (6) in the same way as the variables x, so the system (1) defined on an extended space formed by species concentrations and parameters is still characterized by multi-affine vector fields. The components of the vector fields corresponding to parameters will be zero, meaning that the kinetic constants are assumed constant but unknown within given ranges.*

Remark 3.5 *Algorithm 1 requires the vector field f to be evaluated at each vertex. If there are N_r rectangular sets in the partition, the number of evaluations is $N_r \times 2^N$. If each coordinate is divided into K intervals, $N_r = K^N$. Thus, the number of computations scales as $(2K)^N$. While the complexity of this algorithm is exponential in the number of dimension, the alternative which is exhaustive simulation, is impractical.*

4 Determining the feasible set $\mathcal{F}$

In this section we briefly describe the Rapidly Exploring Random Forest of Trees (RRFT) algorithm, a randomized algorithm that can be used to delineate a feasible set $\mathcal{F}$ — a collection of rectangles $I(q_1, \dots, q_N)$, $(q_1, \dots, q_N) \in \mathbf{Q}$ which contain initial conditions that can reach $\mathcal{P}$. Our algorithm leverages the work in randomized motion planning pioneered by the robotics community. We briefly review this work before introducing our algorithm.

4.1 Motion planning

Probabilistic Road Maps (PRMs) can be used to solve the motion problem [44, 30], which involves finding a path from a starting point to a goal point in configuration space. The problem is usually cast in a geometric setting with no kinematics or dynamics. In contrast, Rapidly exploring Random Trees (RRTs) [36] generate random states for dynamic systems directly by working in the space of admissible input functions $u(t) \in \mathcal{U}$. The algorithm (see Figures 2 and 3) constructs a tree $\mathcal{T}_{x^0}$ rooted at initial state x^0, whose vertices are states $x \in X$ and whose edges are inputs $u(t) \in \mathcal{U}$ which cause the system to evolve from one vertex to a connected vertex. The algorithm constructs the tree beginning

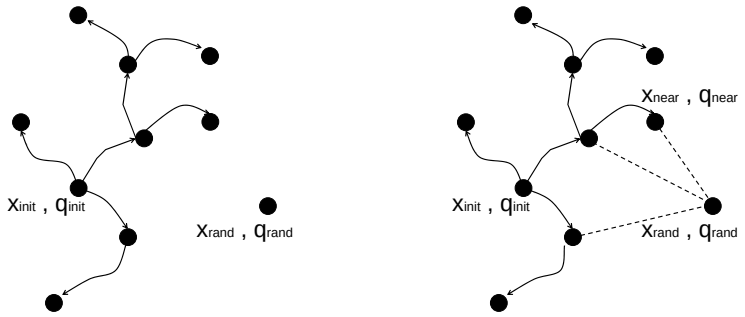

Figure 2: Growth of individual trees in the RRT algorithm [36]). Each tree consists of vertices which are states x, and edges which are input functions $u(t) \in \mathcal{U}$. First a new state is generated at random, x^{rand} (*left*). The algorithm then determines the closest state, x^{near} in the tree to the random state (*right*).

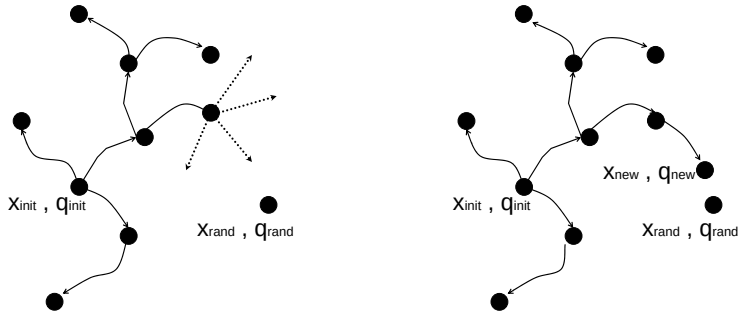

Figure 3: Growth of individual trees in the RRT algorithm (continued from Fig. 2). After finding the closest node, the algorithm determines which $u(t) \in \mathcal{U}$ brings x^{near} closest to x^{rand} (*left*). $u^{new}(t)$ is applied for a pre-determined duration Δt and the new state x^{new} and u^{new} are added to the tree (*right*).

with a user supplied initial state, which we refer to as the seed value. A sample state is generated at random, $x^{rand} \in X$. It is then determined which of the existing states, in the tree are closest to this new state, $x^{near} \in \mathcal{T}_{x^0}$; and which $u^{new}(t) \in \mathcal{U}$, when applied for pre-determined time interval Δt, would bring the system as close as possible to x^{rand}. The resulting new state x^{new} is added as a vertex to $\mathcal{T}_{x^0}$ with $u^{new}(t)$ the input characterizing the edge from x^{near} to x^{new}. This procedure has the effect of growing a tree whose distribution of vertices approaches that of the random distribution which was used to create x^{rand}, causing it to cover the state space rather rapidly. A survey of the algorithm's properties appears in [37].

Of course the application of such algorithms require the ability to simulate system's dynamics. For hybrid systems this requires a special set of algorithms, such as [51] or [19] to properly address the nonsmooth nature; and integration

algorithms capable of handling system evolving at disparate time scales such as [16, 17].

4.2 The RRFT algorithm

Unlike motion planning in which the end goal is to physically steer the system, our intention is merely to determine if it is possible for the system to reach $\mathcal{P}$ from some $x^0 \in I(q_1, \ldots, q_N) \subset \mathbf{X^0}$ within a finite time span $t \in [t_0, t_f]$. If so, the set $I(q_1, \ldots q_N)$ is added to $\mathcal{F}$. All possible $I(q_1, \ldots q_N)$ in $\mathbf{X^0}$ are tested. In this way our usage of the RRT method for analysis rather than synthesis [27, 33, 20] is closely related to our work on test generation for hybrid systems [32, 18]. While the RRT algorithm is in many ways suited to applications such as ours where the system dynamics are complex and high-dimensional, the RRT only addresses time varying inputs such as $u(t)$. Recall that the evolution of our hybrid system is characterized by two elements.

- The initial condition $x^0 \in \mathbf{X^0}$ for the evolution of the state.
- The exogenous modelling noise $\nu(t) \in \mathcal{N}$ that "steers" the system.

In our algorithm, the repeated application of the RRT algorithm results in a tree for every choice of initial condition x^0. Accordingly, we need to consider a *set of trees* that rapidly explore the state space.

One key component of this approach is that each RRT can be computed in parallel on a different CPU's, therefore we assume a fixed computational resource that will dictate the number of trees that can be simultaneously computed in parallel. Let this number be n_t. We propose the Rapidly exploring Random Forest of Trees (RRFT) algorithm as follows. For each set $I(q_1, \ldots, q_N)$ in $\mathbf{X^0}$, a set of seed values $S = \{s^1, \ldots s^{n_t}\}$ is generated from a quasi-random sequence, where each $s^i \in I(q_1, \ldots, q_N)$. RRT's, $\mathcal{T}_{s^1}, \ldots \mathcal{T}_{s^{n_t}}$ are planted for each of these "seed" values. As the RRT algorithm progresses, we monitor the progress of each tree. If at any point the growth of one of the trees (as measured by a function $g(\mathcal{T}_{s^i})$) drops below a threshold $\bar{g}$; or, the coverage of the state space (as measured by some function $c(\mathcal{T}_{s^i})$) exceeds a threshold $\bar{c}$, the tree is terminated. Provided the set $I(q_1, \ldots, q_N)$ is not adequately covered with seeds (again as measured by some function $\mu(S)$) a new "seed" is planted and a new tree is initiated. The process of planting and growing new trees continues until a trajectory linking $I(q_1, \ldots q_N)$ and $\mathcal{P}$ is discovered (in which case $I(q_1, \ldots q_N)$ is added to $\mathcal{F}$), or until $I(q_1, \ldots q_N)$ is sufficiently covered ($\mu(S) \leq \bar{\mu}$) with seed values, whose trees have stopped growing. A new set $I(q'_1, \ldots, q'_N)$ is selected and the process is repeated until all of $\mathbf{X^0}$ has been tested.

We defer the discussion of how to compute the functions $g(\mathcal{T}_{s^i})$, $c(\mathcal{T}_{s^i})$, and $\mu(S)$ until Section 4.3. Note that in the description of the algorithm below, we will use the notation $x^0 + \int^{\Delta t} HS(\nu(t))dt$ to denote the simulation of the hybrid system HS, characterizing the biomolecular network, over a interval time Δt, using the disturbance function $\nu(t)$, and initial condition x^0.

Algorithm 3 Construct a feasible set F

for each $I(q_1, \ldots, q_N)$ in $\mathbf{X^0}$ **do**
  Generate initial seed set $S = \{s^1, \ldots, s^{n_t}\}$ where $s^i \in I(q_1, \ldots, q_N)$
  for $i = 1, \ldots n_t$ **do**
    Initialize RRT's $\mathcal{T}_{s^i}$
  end for
  while (true) **do**
    for $i = 1, \ldots n_t$ **do**
      Extend($\mathcal{T}_{s^i}$)
      if $\mathcal{T}_{s^i} \cap \mathcal{P} \neq \emptyset$ **then**
        add $I(q_1, \ldots, q_N)$ to $\mathcal{F}$
        break
      else
        if $g(\mathcal{T}_{s^i}) \leq \bar{g}$, OR, $c(\mathcal{T}_{s^i}) \leq \bar{c}$ **then**
          terminate $\mathcal{T}_{s^i}$
          $n_t \leftarrow n_t - 1$
          if $\mu(S) > \bar{\mu}$ **then**
            generate new seed point s^{new} and append to S
            initialize $\mathcal{T}_{s^{new}}$
            $n_t \leftarrow n_t + 1$
          end if
        end if
      end if
    end for
    if $n_t = 0$ **then**
      break
    end if
  end while
end for

Algorithm 4 Extend($\mathcal{T}$)

$x^{rand} \leftarrow$ random()
$x^{near} \leftarrow$ nearestNeighbor($\mathcal{T}, x^{rand}$)
$\nu^{new} = \arg\min_{\nu \in \mathcal{N}} \{dist((x^{rand}, x^{near} + \int^{\Delta t} HS(\nu(t))dt)\}$
$x^{new} = x^{near} + \int^{\Delta t} HS(\nu^{new}(t))dt$
add vertex x^{new} to $\mathcal{T}$
add edge ν^{new}, from x^{near} to x^{new}, to $\mathcal{T}$

4.3 Adequacy Criteria

Theoretical results on RRT's from the motion planning literature [37] suggest that as the number of nodes in the tree goes to infinity, the tree should cover the entire reachable set, although it is impossible to determine the reachable set in advance. However, because the input function space must be discretized and because the algorithm is inherently greedy, it is possible for the tree to create new nodes that are very close to the existing nodes. Therefore, there are two plausible reasons to stop growing a tree $\mathcal{T}_{s^i}$: (1) the state space is sufficiently covered that one can be confident no trajectory exists linking $I(q_1, \ldots, q_N)$ and $\mathcal{P}$; or (2) the tree is no longer actively growing.

In order to determine how to allocate our computational resources effectively we must monitor the progress of each tree. In particular, we are interested in three measures: the growth rate of an individual RRT, $g(\mathcal{T}_{s^i})$; the coverage of the state space by an individual RRT $c(\mathcal{T}_{s^i})$; and coverage of the set of possible seeds $I(q_1, \ldots, q_N)$ by S, $\mu(S)$. We explored different measures of growth and coverage including the discrepancy and dispersion [9]; the size of the Vornoi regions [37]; as well as the volume of the convex hull of the tree nodes as well as other bounding polygons. However, we found these measures to be either too expensive to compute in high dimensions or overly conservative. Instead, we begin by overlaying a grid on the state space. Note that this grid *is not* used to construct the tree, merely to assess its coverage and growth. We calculate the minimum distance from each grid point to the set of nodes in the tree. The distance may be thought of as the radius of the largest ball centered at each grid point which does not contain a tree node or another grid point (see Figure 4). Clearly, the maximum value of the radius is δ, the spacing between adjacent grid points. It should be stressed that this list of distances can be updated incrementally as new tree nodes are added, since the affect of each new node is local. We define the coverage of the tree $\mathcal{T}_{s^i}$, $c(\mathcal{T}_{s^i})$, as the average of all the distances obtained in this manner, normalized by the grid spacing.

$$c(\mathcal{T}_{s^i}) = \frac{1}{\delta} \sum_{j=1}^{n_g} \frac{\min(d_j, \delta)}{n_g} \tag{11}$$

where n_g is the number of grid points, and d_i is the radius of the largest ball centered at each grid point. Clearly this measure is a monotonically decreasing function. If it goes to zero on a given grid it tells us that any set whose distance along its smallest dimension is greater than the grid spacing has been entered. Said another way, the state space is covered up to a resolution equal to the grid spacing. This measure is similar to an approximation of dispersion [9], but less conservative and faster to compute.

The derivative of $c(\mathcal{T}_i)$ indicates the growth of the tree. Therefore

$$g(\mathcal{T}_{s^i}) = dc(\mathcal{T}_{s^i})/dn_{v,i} \approx \Delta c(\mathcal{T}_{s^i})/\Delta n_{v,i}, \tag{12}$$

where $n_{v,i}$ is number of vertices in tree rooted at s^i. Overall one of the advantages of this measure is that the grid size can be as fine or coarse as one chooses.

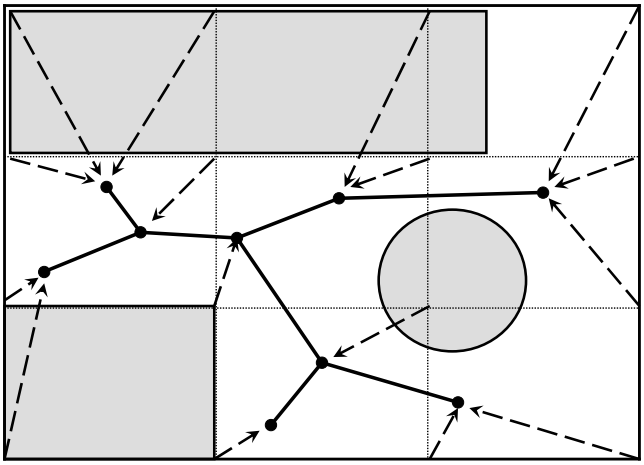

Figure 4: A grid is super imposed on the state space. The shaded regions indicate unreachable sets. The average of the distances from the grid points to the closest nodes (shown as dashed arrows), should converge to a finite number as the tree fills the reachable space.

Finer grids will require more distance queries but are more accurate indications of coverage.

Regarding the coverage of the set of potential seeds $I(q_1, \ldots, q_N)$ by S, one appropriate measure, which first appeared in the Monte Carlo literature and later has been used in the context of RRT's [9], is dispersion. Dispersion is considered the radius of the largest empty ball whose center lies in a set which does not include a sampled point in the set. It is a measure of the largest region which is not covered. Therefore, we use normalized dispersion as a criteria for coverage of $I(q_1, \ldots, q_N)$

$$\mu(S) = \sup_{x \in I(q_1,\ldots,q_N)} \min_{s \in S} d(x, s)/\mu(S)_{max}. \tag{13}$$

The grid based method introduced above can be use to approximate this quantity as well. Fortunately there exists sequences of quasi-random numbers which have low dispersion. Accordingly, we use Halton sequences [23] to generate the seed values.

5 Case study: luminescence control in *Vibrio fischeri*

In this section, we show how the theory and algorithms developed in this paper can be applied to study the behavior of a specific genetic regulatory network. We consider for illustration the phenomenon of bioluminescence production in

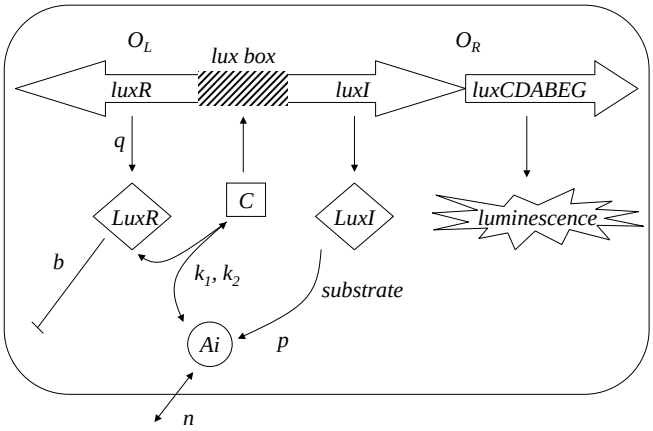

Figure 5: Schematic representation of the genetic network regulating the luminescence production in the marine bacterium *V. fischeri*.

the marine bacterium *V. fischeri*, which is controlled by the transcriptional activation of the *lux* genes [26, 8]. The *lux* regulon is organized in two transcriptional units, O_L and O_R, separated by a regulatory region called the *lux box*, as shown in Figure 5. The leftward operon, O_L, contains the [1] *luxR* gene encoding protein LuxR, a transcriptional regulator of the system. The rightward operon O_R consists of seven genes *luxICDABEG*. The expression of the *luxI* gene results in the production of protein LuxI, which is required for endogenous production of *autoinducer*, *Ai*, a small membrane-permeant signal molecule. The other genes in O_R are involved in the production of luminescence. Finally, the autoinducer *Ai* binds to protein LuxR to form a complex *C*, which has an electronic affinity to the *lux box*. The transcription of both *luxICDABEG* and *luxR* is activated by the binding of *C* to the *lux box*, which is modeled using a continuous piecewise linear activation function (see Figure 6).

A 9-dimensional model for this network is presented in [8]. For illustrative purposes, we consider a simplification that is possible under the assumption that the dynamics of protein LuxI are fast [26]. With this simplification, the system becomes three dimensional ($N = 3$) with state $x = [x_1\, x_2\, x_3]^T$, where x_1, x_2, and x_3 represent the concentrations of protein LuxR, complex C, and autonducer *Ai*, respectively. The main reason for chosing this simplified model is because the reduction in dimensionality allows us to include three-dimensional trajectories and reachability graphs, which would not be possible in higher dimensional space. Examples of analysis in higher dimensional space are presented in [26, 6].

Two additive exogeneous inputs $\nu = [\nu_1\, \nu_2]^T$ ($m = 2$) are present in the model. In the presence of a plasmid that produces LuxR independently, ν_1 is the rate of transcription of the plasmid, while ν_2 models an external source of autoinducer. More generally, they can represent stochastic uncertainty that

[1] We use italics (*e.g.*, *luxR*) to indicate the genes and plain font to denote the protein expressed by the gene (*e.g.*, LuxR)

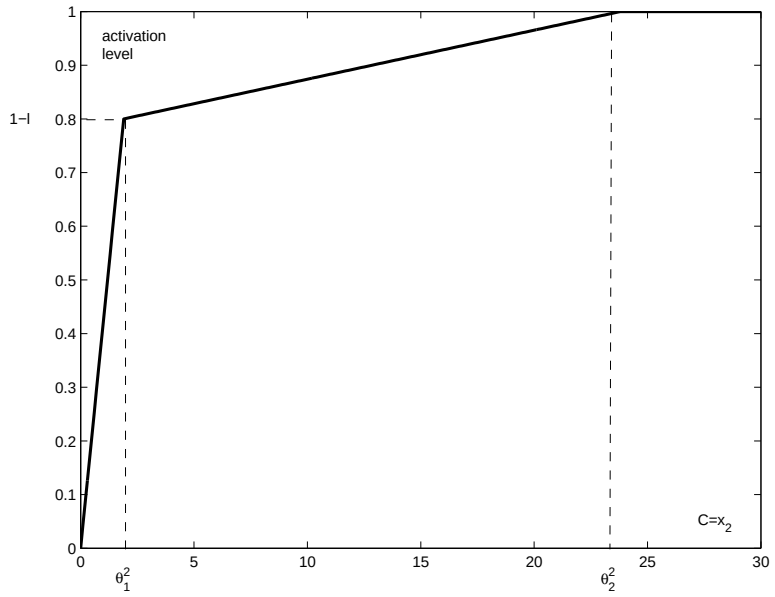

Figure 6: Continous piecewise linear activation function of the lux genes.

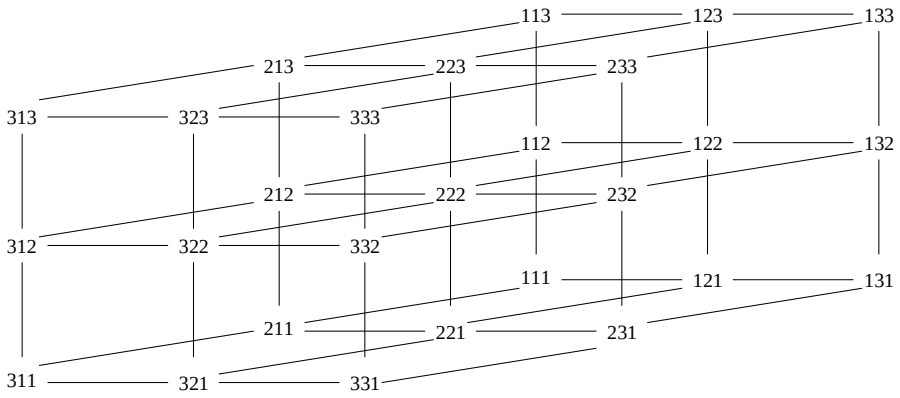

Figure 7: The simple graph for the partitioning in Equation (14).

may be inherent in the model. We will let $\mathcal{N}$ be a rectangular set given by $[0, \nu_{1,max}] \times [0, \nu_{2,max}]$.

Regarding the rectangular partition of the state space, we consider $n_1 = n_2 = n_3 = 3$. The thresholds θ_j^2, $j = 1, 2$ represent the values of x_2 for which the dynamics are changed due to different activation rates (see Figure 6), while $j = 0, 3$ represent physical lower and upper bounds. The other division points θ_j^i, $i = 1, 3$, $j = 0, 1, 2, 3$ were chosen so that the state space is divided into regions of interest, which could be though of "small", "medium" and "large" with respect to the corresponding specie concentrations. The numerical values of these constants are given by by

$$\begin{array}{llll} \theta_1^0 = 0 & \theta_1^1 = 10 & \theta_1^2 = 50 & \theta_1^3 = 100 \\ \theta_2^0 = 0 & \theta_2^1 = 1.9 & \theta_2^2 = 23.8 & \theta_2^3 = 100 \\ \theta_3^0 = 0 & \theta_3^1 = 10 & \theta_3^2 = 50 & \theta_3^3 = 100 \\ \nu_{1,max} = 200 & \nu_{2,max} = 200 & & \end{array} \tag{14}$$

and, by Equation (4),

$$a^{(q_1 q_2 q_3)} = (\theta_1^{q_1 - 1}, \theta_2^{q_2 - 1}, \theta_3^{q_3 - 1}),\ b^{(q_1 q_2 q_3)} = (\theta_1^{q_1}, \theta_2^{q_2}, \theta_3^{q_3}),\ q_{1,2,3} \in \{1, 2, 3\}$$

Following the notation introduced in Section 2, the system can be represented as the simple graph in Figure 7. The dynamics in each of the 27 rectangles $R_3(a^{(q_1 q_2 q_3)}, b_{(q_1 q_2 q_3)}) = I(q_1 q_2 q_3)$ $q_{1,2,3} \in \{1, 2, 3\}$ are given by:

$$\dot{x} = f^{(q_1 q_2 q_3)}(x) + B\nu \tag{15}$$

where

$$f^{(q_1 q_2 q_3)} = \begin{bmatrix} k_2 x_2 - k_1 x_1 x_3 - b x_1 + q r^{(q_1 q_2 q_3)} \\ k_1 x_1 x_3 - k_2 x_2 \\ k_2 x_2 - k_1 x_1 x_3 - n x_3 + p r^{(q_1 q_2 q_3)} \end{bmatrix}, \quad B = \begin{bmatrix} s & 0 \\ 0 & 0 \\ 0 & n \end{bmatrix} \tag{16}$$

and

$$r^{(q_1 1 q_3)} = \frac{(1 - l) x_2}{\theta_2^1},\ r^{(q_1 2 q_3)} = 1 - \frac{l(\theta_2^2 - x_2)}{\theta_2^2 - \theta_2^1},\ r^{(q_1 3 q_3)} = 1,$$

for q_1, $q_2 = 1, 2, 3$ and $l = 0.2$ (see Figure 6). The dynamics are everywhere continuous, and, therefore, the vector fields on adjacent rectangles coincide on the common facet. The significance of the state variables and parameters is given in the following table:

x_1	=	protein LuxR (ml^{-3})
x_2	=	complex C (ml^{-3})
x_3	=	autoinducer $Ai(ml^{-3})$
k_1	=	binding rate constant ($30\ l^3 m^{-1} t^{-1}$)
k_2	=	dissociation rate constant ($10\ t^{-1}$)
n	=	diffusion constant ($10\ t^{-1}$)
b	=	degradation constant for LuxR ($3t^{-1}$)
p	=	formation of Ai due to lux gene activity ($30ml^{-3}t^{-1}$)
q	=	formation of LuxR due to lux gene activity ($5ml^{-3}t^{-1}$)
s	=	scaling constant ($10\ t^{-1}$)

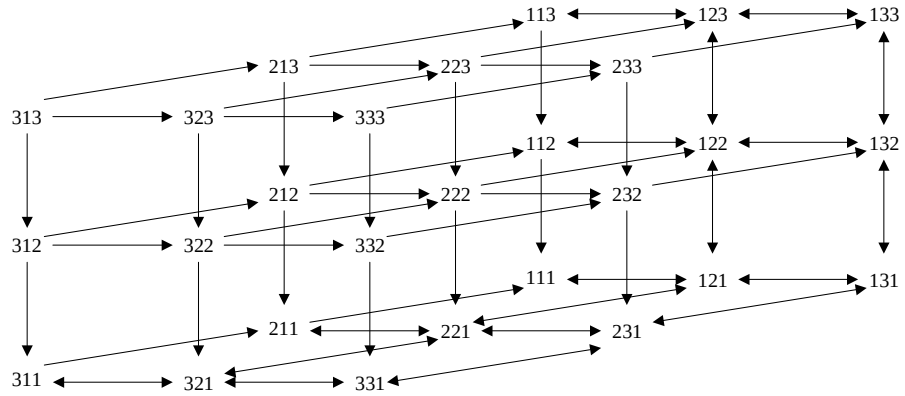

Figure 8: The oriented graph obtained by applying Algorithm 1 to the simple graph in Figure 7.

5.1 Obtaining the infeasible set $\mathcal{I}$ by using MARP

Since the vector field given by (15), (16) with $\nu_1 = \nu_2 = 0$ is continuous, we can simply apply Algorithm 1 to determine the orientation of the edges of the graph given in Figure 7. The result is given in Figure 8.

Assume the property set $\mathcal{P}$ is given by the target rectangle (222). This requires the execution of the **repeat** loop in Algorithm 2 four times:

- $\mathcal{R} := \mathcal{P} = \{(222)\}$
- $\mathcal{R} = \{(222), (223), (322), (212)\}$
- $\mathcal{R} = \{(222), (223), (322), (212), (213), (323), (312)\}$
- $\mathcal{R} = \{(222), (223), (322), (212), (213), (323), (312), (313)\}$

The infeasible set $\mathcal{I}$ is the complement of $\mathcal{R}$, and it consists of the remaining $27-8 = 19$ rectangles. The biological significance of this result is that luminescence (which is described by relatively high values of complex x_2 and autoinducer x_3) can only be achieved if the initial level of autoinducer x_3 is high.

5.2 Obtaining a feasible set $\mathcal{F}$ using the RRFT algorithm

In this section, we consider the construction of feasible set $\mathcal{F}$ for the system with noise $\nu(t)$ using RRFT. In Figure 9, several sample trajectories illustrate the computation of candidate trajectories corresponding to the (111) to (121) transition, and the (121) to (111) transition. Each of these trajectories is the result of planting a tree with a different seed. Thus the RRFT algorithm can be used to obtain sample trajectories for edges in the directed graph in Figure 8.

Figure 10 illustrates the case when no solution can be found. We consider the property set (313) and consider initial values from the rectange (111). Each

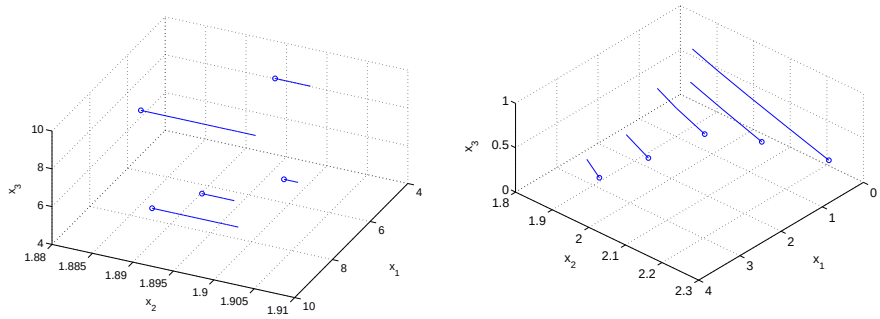

Figure 9: Sample trajectories from 111 to 121 (left) and 121 to 111 (right).

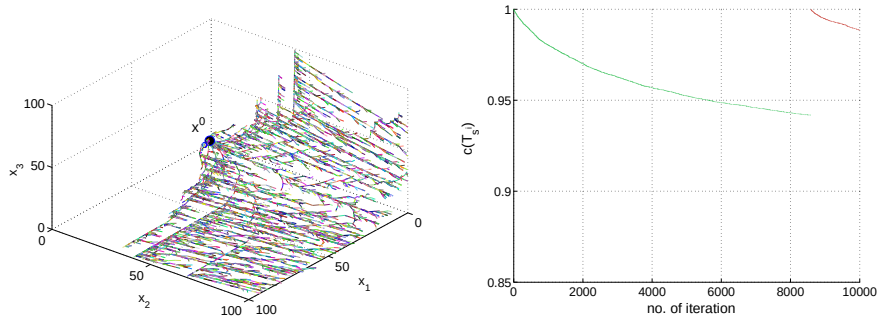

Figure 10: A tree with the specified initial condition, x^0, from the (111) rectangle (left) ($\Delta t = 0.001$). The coverage $c(\mathcal{T}_{s^i})$ for the tree (right). A new tree is started when the growth rate slows below a specified threshold ($\bar{g} = 1 \times 10^{-6}$ used in this example).

tree is grown until the growth rate is unacceptably low upon which a new tree is seeded. The growth of a tree (left) and the converage of the tree $c(\mathcal{T}_{s^i})$ (right) are shown in the figure. As new vertices are generated, $c(\mathcal{T}_{s^i})$ and $g(\mathcal{T}_{s^i})$ decrease (right). After reaching a specified threshold, the algorithm terminates the tree and generates a new tree from a new seed. Figure 11 shows the coverage of the seed set as new seeds are generated. Seeds are generated using Halton sequence.

Recall that the property set P is the rectangle (222). We first consider the rectangle (111) to determine candidate points for the feasible set $\mathcal{F}$. Figure 12 shows the forest of trees where a solution trajectory is found. Ten initial seeds are generated and a forest starts to grow until a solution is found. The solution trajectory, the modes and the transitions are shown in Figure 13. Figure 14 shows $c(\mathcal{T}_{s^i})$ for the tree corresponding to a solution. No new seeds are generated beyond the initial set in this case. One of the ten trees is able to find the trajectory.

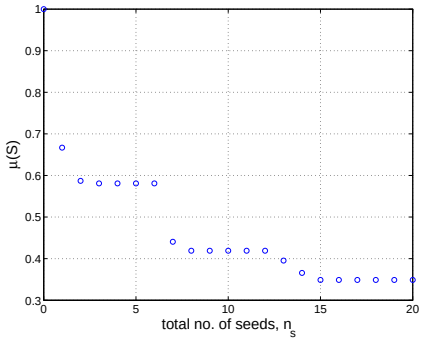

Figure 11: The coverage improves ($\mu(S)$ decreases) as new trees are seeded.

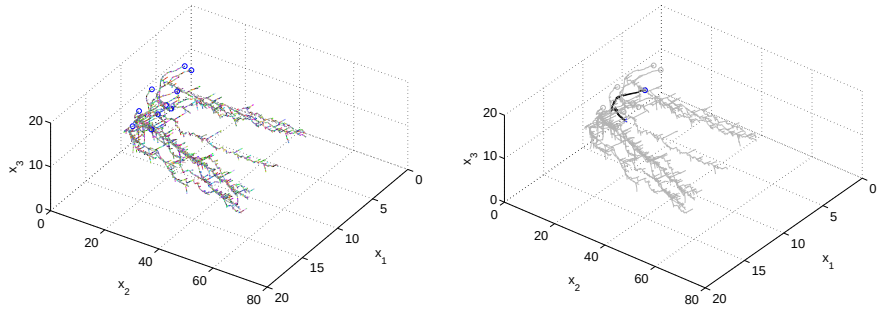

Figure 12: The forest of trees computed by Algorithm 3 with initial conditions from the (111) rectangle with the goal of finding a trajectory that reaches (222). The forest is shown on the left. The trajectory found by the algorithm that reaches (222) is highlighted (shown dark) on the right ($\Delta t = 0.0001$).

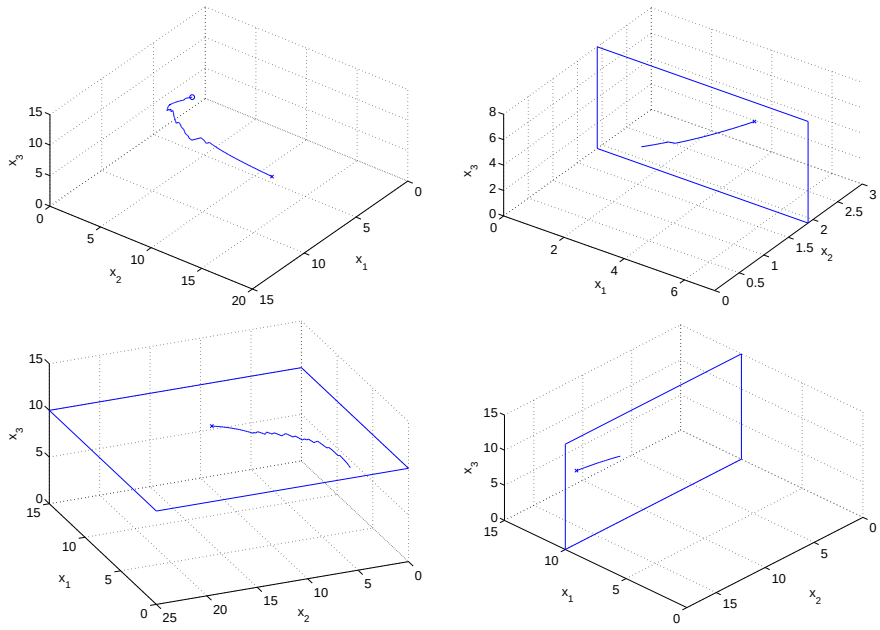

Figure 13: The trajectory found by Algorithm 3, showing the different modes. (a) Top left - Solution trajectory; (b) Top right - (111) to (121); (c) Bottom left - (121) to (122); (d) Bottom right - (122) to (222).

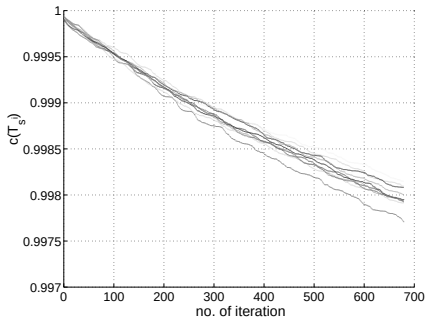

Figure 14: Coverage $c(\mathcal{T}_{s^i})$ for each tree. A solution trajectory is found among the initial 10 seeds.

6 Conclusion

Hybrid systems are widely used in robotics to model the use of specific controllers and estimators in different regimes, switches based on contact mechanics of rolling or sliding, and also to take into account the interactions and messaging in a multi-robot team. Hybrid systems also arise naturally as models of genetic and metabolic networks. They capture the switching behavior that is observed in phenomena such as transcription, protein-protein interactions, and cell division and growth and also to provide global descriptions of biological systems described locally around operating points. In this paper, we develop computationally efficient techniques to analyze hybrid models of bio-molecular networks by exploiting their specific structure. We defined the framework of multi-affine rectangular hybrid systems, where the vector fields have product type nonlinearities to capture the dynamics of chemical reactions and the invariants are rectangular, because different behaviors emerge as a function of different ranges of concentrations of regulatory species. To prove qualitative properties of such systems, which are biologically significant, we developed the Multi-Affine Rectangular Partition (MARP) algorithm and the Rapidly Exploring Random Forest of Trees (RRFT) algorithm, and illustrate their usefulness by using as a case study, the phenomenon of luminescence production in the marine bacterium *V. fischeri*. While the case study was deliberately chosen to enable a low-dimensional model to facilitate graphical illustration through plots, the techniques here are applicable to very high-dimensional systems [3]. Future work is being directed towards developing tools for formal analysis of larger classes of hybrid systems, which could capture more complicated biochemical phenomena, and developing control laws for species in the network that can be directly controlled from outside the cell.

References

[1] Thomas A.Henzinger, Benjamin Horowitz, Rupak Majumdar, and Howard Wong-Toi. Beyond HyTech: Hybrid systems analysis using interval numerical methods. In *Hybrid Systems : Computation and Control*, volume 1790 of *Lecture Notes in Computer Science*, pages 130–144. Springer Verlag, 2000.

[2] R. Alur, C. Belta, F. Ivancic, V. Kumar, M. Mintz, G. Pappas, and J. Schug. Hybrid modelling and simulation of biomolecular networks. In *Lecture Notes in Computer Science*, volume 2034, pages 19–32. 2001.

[3] R. Alur, C. Belta, F. Ivancic, V. Kumar, H. Rubin, J. Schug, O. Sokolsky, and J. Webb. Visual programming for modeling and simulation of bioregulatory networks. In *International Conference on High Performance Computing*, Bangalore, India, 2002.

[4] R. Alur, C. Belta, V. Kumar, M. Mintz, G. J. Pappas, H. Rubin, and J. Schug. Modeling and analyzing biomolecular networks. *Computing in Science and Engineering*, pages 20–30, Jan/Feb 2002.

[5] Eugene Asarin, Thao Dang, and Oded Maler. d/dt: A tool for reachability analysis of continuous and hybrid systems. In *5th IFAC Symposium on Nonlinear Control Systems*, St. Petersburg, Russia, 2001.

[6] C. Belta, P. Finin, L.C.G.J.M. Habets, A. Halasz, M. Imielinksi, V.Kumar, and H. Rubin. Understanding the bacterial stringent response using reachability analysis of hybrid systems. submitted to HSCC 2004.

[7] C. Belta, L. Habets, and V. Kumar. Control of multi-affine systems on rectangles with applications to hybrid biomolecular networks. In *41st IEEE Conference on Decision and Control*, Los Angeles, NV, 2002.

[8] C. Belta, J. Schug, T. Dang, V. Kumar, G. J. Pappas, H. Rubin, and P. V. Dunlap. Stability and reachability analysis of a hybrid model of luminescence in the marine bacterium vibrio fischeri. In *40th IEEE Conference on Decision and Control*, Orlando, FL, 2001.

[9] Michael Branicky, Steven LaValle, Kari Olson, and Libo Yang. Quasi-randomized path planning. In *IEEE International Conference on Robotics and Automation*, pages 1481–1487, Seoul, Korea, May 2001.

[10] D. L. Brutlag, A. R. Galper, and D. H. Millis. Knowledge-based simulation of dna metabolism: prediction of enzyme action. *Computer Applications in the Biosciences*, 7(1):9–19, 1991.

[11] Oleg Butchkarev and Stavros Tripakis. Verification of hyrbid systems with linear differential inclusions using elipsoidal roximations. In *Hybrid Systems : Computation and Control*, volume 1790 of *Lecture Notes in Computer Science*, pages 73–88. Springer Verlag, 2000.

[12] Alongkrit Chutinam and Bruce Krogh. Computational techniques for hybrid system verification. *IEEE transactions on automatic control*, 48(1):64–75, 2003.

[13] Daniel T. Gillespie. Fluctuation and Dissipation in Brownian Motion. *Am. J. Phys*, 61():1077–1083, 1993.

[14] A. Das, R. Fierro, J. Ostrowski V. Kumar, J. Spletzer, and C. J. Taylor. Vision based formation control of multiple robots. In *IEEE Transactions on Robotics and Automation*, volume 18, pages 813–825, 2002.

[15] H. de Jong, J. L. Gouze, C. Hernandez, M. Page, T. Sari, and J. Geiselmann. Hybrid modeling and simulation of genetic regulatory networks: a qualitative approach. In *Hybrid Systems: Computation and Control*, 2003.

[16] J. Esposito and V. Kumar. Efficient dynamic simulation of robotic systems with hierarchy. In *IEEE International Conference on Robotics and Automation*, pages 2818–2823, Seoul, Korea, May 2001.

[17] Joel Esposito, George Papas, and Vijay Kumar. Multi-agent hybrid system simulation. In *IEEE Conference on Decision and Control*, Orlando, FL, December 2001.

[18] Joel M. Esposito, Jongwoo Kim, and Vijay Kumar. A probabilistic approach to automated test case generation for hybrid systems. Submitted to Hybrid Systems: Computation and Control 2004.

[19] Joel M. Esposito, George J. Pappas, and Vijay Kumar. Accurate event detection for simulating hybrid systems. In M.D. Di Benedetto and A. Sangiovanni-Vincentelli, editors, *Hybrid Systems : Computation and Control*, volume 2034 of *Lecture Notes in Computer Science*, pages 204–217. Springer Verlag, 2001.

[20] E. Frazzoli, M. Dahleh, and E. Feron. Real-time motion planning for agile autonomous vehicles. *AIAA Journal of Guidance, Control, and Dynamics*, 25(1):116–129, 2002.

[21] D.T. Gillespie. Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.*, 81:2340–2361, 1977.

[22] L. Glass. Classification of biological networks by their qualitative dynamics. *Journal of Theoretical Biology*, 54:85–107, 1975.

[23] J. H. Halton. On the efficiency of certain quasi-random sequences of points in evaluating multi-dimensional integrals. *Numer. Math.*, (2):84–90, 1960.

[24] R. Heinrich and S. Schuster. *The regulation of cellular systems.* Chapman and Hall, New York, 1996.

[25] Thomas A. Henzinger, Peter W.Kopke, Anuj Pujri, and Pravin Varaiya. What is decidable about hybrid automata? In *Proceedings of the 27th annual ACM Symposia on the Theory of Computation STOC*, pages 373–382, May 1995.

[26] S. James, P. Nilsson, G. James, S. Kjelleberg, and T. Fagerstrom. Luminescence control in the marine bacterium vibrio fischeri: an analysis of the dynamics of lux regulation. *Journal of molecular biology*, 296:1127–1137, 2000.

[27] T. Karatas and F. Bullo. Randomized searches and nonlinear programming in trajectory planning. In *cdc*, pages 5032–5037, Orlando, FL, dec 2001.

[28] S. A. Kauffmann. Metabolic stability and epigenesis in randomly constructed genetic nets. *Journal of Theoretical Biology*, 22:437–467, 1969.

[29] S. A. Kauffmann. *The origins of order: self-organization and selection in evolution*. Oxford University Press, New York, 1993.

[30] Lydia E. Kavraki, P. Svestka, J.C.Latombe, and M.H.Overmars. Probabilistic roadmaps for path planning in high dimensional configuration space. *International Transactions on Robotics and Automation*, 12:566–580, 1996.

[31] Thomas B. Kepler and Timothy C. Elston. Stochasticity in transcriptional regulation: Origins, consequences, and mathematical representations. *Biophysical Journal*, 81:3116–3136, December 2001.

[32] J. Kim, J. Keller, and V. Kumar. Design and verification of controllers for airships. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, Las Vegas, NV, 2003.

[33] J. Kim and J. Ostrowski. Motion planning of aerial robot using rapidly-exploring random trees with dynamic constraints. In *IEEE International Conference on Robotics and Automation*, Taipei, Taiwan, 2003.

[34] B. Kuipers. Qualitative simulation. *Artificial intelligence*, 29:289–388, 1981.

[35] George Lafferriere, George J. Pappas, and Sergio Yovine. A new class of decidable hybrid systems. In *Hybrid Systems : Computation and Control*, volume 1590 of *Lecture Notes in Computer Science*, pages 137–151. Springer Verlag, 1999.

[36] Steven M. LaValle and James J. Kuffner. Randomized kinodynamic planning. *International Journal of Robotics Research*, 20(5):378–400, 2001.

[37] Steven M. LaValle and James J. Kuffner. Rapidly exploring random trees: Progress and prospects. In B. Donald, K. Lynch, and D.Rus, editors, *Algorithmic and computational robotics: new directions*, pages 293–308. A.K. Peters, Wellesley, MA, 2001.

[38] N. Lynch and B. H. Krogh, editors. *Hybrid Systems: Computation and Control*, volume 1790 of *Lecture Notes in Computer Science*. Springer-Verlag, Berlin, 2000.

[39] H. M. McAdams and A. Arkin. Stochastic mechanisms in gene expression. *Proceedings of the National Academy of Science of the USA*, 94:814–819, 1997.

[40] T. Mestl, E. Plathe, and S. W. Omholt. Periodic solutions in systems of piecewise-linear differential equations. *Dynamics and stability of systems*, 10(2):179–193, 1995.

[41] Ian Mitchell. *Application of level set methods to control and reachability problems in continuous and hybrid systems*. PhD thesis, Stanford University, Palo Alto, CA, 2002.

[42] Ian Mitchell and Claire Tomlin. Level set methods for computation in hybrid systems. In B. Krogh and N. Lynch, editors, *Hybrid Systems : Computation and Control*, volume 1790 of *Lecture Notes in Computer Science*, pages 310–323. Springer Verlag, 2000.

[43] E. Mjolsness, D. H. Sharp, and J. Reinitz. A connectionist model of development. *Journal of Theoretical Biology*, 152:429–453, 1991.

[44] N.M.Amato and Y.Wu. A randomized roadmap method for path and manipulation planning. In *IEEE International Conference on Robotics and Automation*, pages 113–120, May 1996.

[45] M. Peschel and W. Mende. *The predator-prey model: do we live in a Volterra world?* Akademie Verlag, Berlin, 1986.

[46] R.Alur and D.Dill. A theory of timed automta. *Theoretical Compuer Science*, 126:183 – 235.

[47] M. A. Savageau. Biochemical systems analysis: I. some mathematical properties of the rate law for the component enzymatic reactions. *J. Theor. Biol*, 25:365–369, 1969.

[48] M. A. Savageau and E. O. Voit. Recasting nonlinear differential equations as s-systems: a canonical nonlinear form. *Math. Biosci.*, 87:83–115, 1987.

[49] P. Tabuada and G.J. Pappas. Model checking LTL over controlable linear systems is decidable. In *Hybrid Systems : Computation and Control*, volume 2623 of *Lecture Notes in Computer Science*. Springer Verlag, 2003.

[50] R. Thomas. Regulatory networks seen as asynchronous automata: a logical description. *Journal of Theoretical Biology*, 153:1–23, 1991.

[51] T.Park and P.Barton. State event location in differential-algebraic models. *ACM transactions on modeling and computer simulation*, 6(2):137–165, 1996.