Algebraic Systems Biology: Theses and Hypotheses*

Bud Mishra[1,2]

[1] Courant Institute, New York University, New York, NY, U.S.A.
[2] NYU School of Medicine, New York University, New York, NY, U.S.A.
mishra@nyu.edu

Abstract. What is systems biology? What can biologists gain from an attempt to algebraize the questions in systems biology? Starting with plausible biological theses, can one algebraically model them and then manipulate them to suggest meaningful hypotheses? Using these hypotheses, can one measure and mine suitable experimental data to validate or refute these hypotheses? Through these intertwined processes of measuring, mining, modeling and manipulating biological systems, can one generate the set of theses and hypotheses upon which systems biology will be founded? This review provides one algorithmic-algebraist's somewhat idiosyncratic response to these and other related questions, but also aims to persuade young algebraists to examine the possible role they and algebra can play to enrich this subject.

1 Hypotheses Non Fingo: Hooke Meets Newton

Over the last few years, Sir Robert Hooke, a somewhat maligned, but still a very fascinating English experimental scientist, had begun to feature unexpectedly prominently in practically all my public presentations on Systems Biology. Initially, what had attracted me to the story of Hooke, was the uncanny resemblance he bore to many contemporary scientists in terms of their insistence on data, observations and hypotheses, their apparent non-rigorous and intuitive approaches to scientific questions, but most inexplicably, their protracted and debilitating open rivalries over the questions of recognition. But, as I learned more about Hooke's life and views, it also became clearer that his indirect influence on the way we think about science today is only surpassed by the opinions of only a handful of other contemporary thinkers, with some of whom Hooke fought bitter and hopeless semi-philosophical battles. They have, thus, unwittingly lent us a useful perspective that is worth examining with some care. How the emerging field of systems biology could establish itself, how it should face its trials and tribulations along the way, and how it could be a significant component of the "new new" biology, etc., could all be examined from the points of view of these 17th century scientists—a viewpoint that remains anachronically and peculiarly relevant even today.

Robert Hooke (1635-1703) was an experimental scientist, mathematician, architect, and astronomer. He was also the first Secretary of the Royal Society from 1677 to 1682, and because of his wide ranging interests, Hooke has been variously described as the "England's Da Vinci." His work Micrographia of 1665 contained his microscopical

* The work reported in this paper was supported by two grants from NSF ITR program.

H. Anai, K. Horimoto, and T. Kutsia (Eds.): AB 2007, LNCS 4545, pp. 1–14, 2007.

investigations, which included the first identification of biological cells, an enduring discovery that has maintained its central place in subsequent developments in biology for more than three centuries. In his drafts of Book II, Newton had referred to him as the most illustrious Hooke—"Cl[arissimus] Hookius." However, not long after, Hooke became involved in a bitter dispute with Sir Isaac Newton over the priority of the discovery of the inverse square law of gravitation. In a letter Hooke wrote to Halley, he complained about omission of credit given to his discovery of the properties of gravity, "which of late Mr. Newton has done me the favour to print and publish as his own inventions." In response Newton wrote back to Halley, "Now is this not very fine? Mathematicians that find out, settle & do all the business must content themselves with being nothing but dry calculators & drudges I beleive[sic] you would think him a man of a strange unsociable temper"—perhaps still a common protest of many unhappy mathematicians whose contributions have been ignored or forgotten. In a more well-known letter that Newton wrote directly to Hooke, he famously said, "If I have seen further[sic] than other men, it is because I have stood on the shoulders of giants"—where, of course, the giants Newton was alluding to were Kepler and Galileo, and not the dwarfish, small-minded and short-tempered likes of Hooke! When Christopher Wren was brought in to resolve this rather strangely English war-of-words, Wren diplomatically described the disagreement using Clairaut's characterization of "the great distance between a glimpsed truth and a demonstrated truth"—raising perhaps, the question of relative roles that should be ascribed to the inductive hypothesis-driven science with respect to the deductive principle-driven science—theses vs. hypotheses.

What is the nature of "TRUTH" in biology, and how is it to be sought? Hooke saw biology as an observational science; he wrote in Micrographia, "The truth is, the science of Nature has already been too long made only a work of the brain and the fancy. It is now high time that it should return to the plainness and soundness of observations on material and obvious things," —a view supporting hypothesis-driven experimentation that advances science through steps of falsification or validation. Newton, on the other hand, championed a search for deep and unifying principles. Newton shunned hypotheses; his motto stated in Principia was "Hypotheses non fingo." ("I feign no hypotheses.") Newton's viewpoints are probably best stated by his most ardent disciple, Halley; in his rather ornately titled essay 'The true Theory of the Tides, extracted from that admired Treatise of Mr. Issac Newton, Intituled, Philosophiae Naturalis Principia Mathematica,' he wrote the following: "Truth being uniform and always the same, it is admirable to observe how easily we are enabled to make out very abstruse and difficult matters, when once true and genuine Principles are obtained."

Biology still remains an observational science; it continues to move through the toils of a vast army of scientists each examining a small subsystem of a favored organism, as the scientists sharpen their intuitions, build upon guesses, conjectures, and hypotheses, and refine their ideas in many small steps—occasionally interrupted by a great leap, a grand vision or a comprehensive shift in paradigm. If subtle principles are to be brought to light, they must wait for serendipity. It has been argued that life is complex, it does not yield to few small neat explanations or pigeon-holing, and if there is a unifying principle in biology, it is that there is no unifying principle in biology.

Can ideas from algorithms and algebra be brought to bear to systematically hunt for principles and patterns that will reveal a grand unified theory of biology? Are their design rules at play in how these systems evolve, interact, and self-assemble? What algebraic tools must we build, if we wish to create a global view of biology? What can be automated to make computers work on tasks that are humanly impossible? Is algebraic systems biology the answer to the problems of biology?

2 Systems Biological Models

2.1 Processes

We start with the following taxonomy into which the cellular biochemical processes are typically organized, as described below.

GENETIC REGULATION: The oft-repeated "central dogma of biology" states that biochemical information in cells is encoded primarily in the Deoxyribo Nucleic Acid (DNA) molecules. DNA is *transcribed* into messenger Ribo Nucleic Acid (mRNA), and the mRNA then is *translated* into proteins at the ribosomes. Genetic regulation is the process of modulation of the expression of the relevant genes at the correct locations and times, and is keyed by specific proteins called transcriptional factors. Through transcriptional factors and other ancillary modulators, proteins, the products of genes, themselves partake in this genetic regulatory process, thus giving rise to complex interaction networks; such proteins interact with regions of the DNA to effect modulation of how genes are transcribed. The binding of the transcription machinery and the transcriptional factors to the DNA involves complex protein-DNA-protein interactions, where, more often than not, the structural modification of the DNA (such as euchromatin and heterochromatin regions) and the protein has to be accounted for.

The rate of gene transcription, the post-transcriptional mechanisms that affect mRNA half-life (i.e., stability) and the formation of the mRNA-ribosome complex are other aspects of genetic regulation. Similarly, there are post-translational mechanisms for protein modification such as phosphorylation of key residues, multimerization, chaperone-guided complex formation, protein-folding control, and genetic control by small interfering RNA (siRNA).

SIGNAL TRANSDUCTION: The cell responds to external signals through receptors, which may be on its surface or in its cytoplasm. The signal is transmitted to the interior through messengers, which induce the desired response to the external signal. Typically, a ligand binds to a trans-membrane receptor whose conformation subsequently changes. This change is detected by proteins bound to it (usually on the cytoplasmic side), or is manifested as a change in the receptor's chemical properties. Subsequently, second messenger molecules amplify the signal and communicate it to the target(s). Alternatively, the ligand can directly enter the cell through non-specific channels and then bind to the receptors inside the cell. Small molecules like calcium often participate in these pathways, where most of the reactants are enzymatic proteins. The net result of the signal transduction pathway is an appropriate response by the specific subcellular component. Very often, the signaling pathway results in the nuclear localization of

transcription factors, leading to the transcription (or shutting down) of corresponding genes. The binding of the signaling molecule with the receptor, the modification of the structure of the receptor and associated proteins (with the receptor sometimes acting as an enzyme) and dispatching of second messengers are the activities near the cell membrane. Receptor desensitization, internalization and regeneration are other complex sub-processes, thus altering the physical properties of binding and diffusion.

METABOLISM: Metabolism represents almost all processes that are not genetic regulatory or signal transducing. The gigantic set of biochemicals needed by the cell are continuously produced and consumed by complex enzyme catalyzed pathways. These comprise the metabolic network. They essentially govern the matter and energy cycles of a cell— the way energy and matter are obtained, transformed and consumed by living organisms. Photosynthesis for example is the process by which light energy is converted into chemical energy during sugar (e.g., glucose) formation. During respiration, the oxidation of glucose transforms the energy into Adenosine Tri-Phosphate (ATP). While the ATP-cycle and photosynthesis comprise the well-known energy metabolism, carbohydrate metabolism deals with Glycolysis and Phosphates, lipid metabolism pertains to Triacyl Glycerol and Fatty Acids, and amino acid metabolism mostly refers to Glutamate and Urea.

OTHER PROCESSES: Biology is complex, and of course, there are still more aspects to cellular biology beyond this simple trichotomic characterization. These include the biophysics of DNA packaging, small interfering RNA (siRNA), protein folding and DNA-protein interaction, cell adhesion, non-transcriptional regulatory pathways, cellular compartments and related spatio-temporal phenomena, cell proliferation, and cell migration. While the modeling approaches suggested here, when further augmented with suitable stochastic and spatial formalisms, will generalize as well, I will not emphasize those applications directly in my discussion here.

2.2 Models

Algorithmic algebraic models of biological systems are created through a process of conceptual simplification. Models, created in this fashion, must strike a balance among fidelity, expressivity and ability to be manipulated algorithmically. For this purposes, the different component parts and processes in the biochemical domain may be represented at different levels of abstraction [22,37]. I summarize some of the major approaches below, but will guide the discussion towards hybrid automata representation, a very general and powerful model for these systems.

LOGICAL MODELING: The state of the reactant is captured through a finite number of abstract-states (where intermediate expression levels are assumed to have the same behavior), and functions are used to describe the new states (concentration range) of the chemical species, given their old states. The transitions between states can be assumed to occur synchronously or (more accurately) asynchronously. In the simplest case, only two states ("on" and "off") are used, and Boolean algebra is used to describe the dynamics. Literature on Concurrent Transition Systems [20,19] and Pathway (Rewrite)

Logic [25] provides good expositions of logical modeling. Kappler et al. [38] demonstrate how to extend simple Boolean networks by using ordinary differential equations to capture the concentration, while Boolean functions continue to determine the rates of the reactions. The probability of being in a state is sometimes a more reasonable measure to estimate, as in the case of Sachs et al. [57], who use Bayesian networks to model cell signaling pathways. Similarly, Shmulevich et al. [58] describe the use of probabilistic Boolean networks to model genetic regulatory networks and determine the long-term joint probabilistic behavior of a few selected genes. Platzer et al. [55] simulate the embryonic development of *C. elegans* by assuming Boolean states for the genes and synchronously updating at each time step based on an interaction matrix. Batt et al. [12] have applied model checking theory on biochemical systems modeled though qualitative simulation.

DIFFERENTIAL EQUATIONS: If instead the concentrations are represented exactly in the real continuous domain, the ordinary differential equations (ODEs) of the dynamics directly follow from the law of general mass action (GMA) [21,39,59]. For instance, in the reaction $aA + bB \longleftrightarrow cC + dD$, the rate of the forward reaction $v_f \equiv k_f[A]^a[B]^b$ and the rate of the backward reaction $v_b \equiv k_b[C]^c[D]^d$, where k_f and k_b are the forward and backward rate constants respectively and the rate of individual reactants is $\frac{1}{c}\dot{C} = \frac{1}{d}\dot{D} = -\frac{1}{a}\dot{A} = -\frac{1}{b}\dot{B} = (v_f - v_b)$. As a compromise between discrete and continuous representations, qualitative differential equations can be used with qualitative states corresponding to the different concentration ranges [12,23]. Partial differential equations are necessary for spatially distributed models, e.g., pde's, sde's, or reaction-diffusion equations.

HYBRID SYSTEMS: Many biological systems, such as the cell, follow a combination of discrete and continuous behaviors, which cannot be characterized in a proper way using either only discrete or only continuous models. On one hand, their evolution is ruled by a continuous dynamical law concerning substance concentrations and gradients, and, on the other hand, such a dynamical law may change discretely depending on the system status itself. Because of their hybrid nature, part discrete and part continuous, such systems are named hybrid systems. To model hybrid systems, Alur et al. introduced the notion of hybrid automata in [3]. Intuitively a hybrid automaton is a "finite-state" automaton with continuous variables, which evolve according to a set of continuous laws characterizing each discrete mode of the automaton itself. The use of hybrid automata for modeling biomolecular networks has been described by Alur et al. [1] and Mishra et al. [46]. Amonlirdviman et al. [7] demonstrated the utility of hybrid systems by modeling Drosophila planar cell polarity. Starting with the S-System formulation of Savageau and Voit [60], Antoniotti et al. [11] used an additional automaton to broaden the set of representable systems, subsequently using full-fledged hybrid automata [10]. Ghosh et al. presented both delta-notch [29,28] and protein signaling network [30] models based on the hybrid automaton formalism. Casagrande et al. [16] suggested a simple (and decidable) hybrid automaton model for the *E. coli* chemotaxis. Lincoln and Tiwari [43] detail hybrid automaton modeling of biochemical networks, while Hu et al. [36] describe stochastic hybrid system modeling of subtilin production in *Bacillus subtilis*. More recently, Drulhe et al. [24] have described piecewise-affine models of genetic regulatory networks.

ALGEBRAIC HYBRID AUTOMATA, TEMPORAL LOGIC AND ALGORITHMS: To create a comprehensive theoretical framework for systems biology, what is needed is an appropriate generalization of discrete-time systems, classical temporal logic, possible-world models of temporal logic given by Kripke (e.g., Kripke structures), model checking algorithms based on graph theoretic analysis, etc. to this richer and more powerful domain. However, the generalization must be suitably powerful to capture reasoning processes closely resembling what is used by the biologists, and yet it should also be appropriately constrained so that these systems can be reasoned by feasible computational means. At the least, the resulting problems should be decidable (computable). We seek such a framework below by a judicious amalgamation of symbolic algebra (using decision procedures of semi-algebraic geometry), sufficiently constrained dense-time logic and algebraic models based hybrid automata. We start with a discussion of such hybrid automata and their reachability problem.

3 Algebraic Systems Biological Models

The subject *Algorithmic Algebraic Model Checking* was introduced to examine connections between systems biology, dynamical systems, modal logic and computability, and how they can be useful in the biological context. Towards this aim, one could begin by addressing the symbolic model checking problem for a new class of hybrid models arising in systems biology – *semi-algebraic hybrid systems*, introduced in the first paper of our "AAMC" (Algorithmic Algebraic Model Checking) series [53]. There, our goal was to characterize the widest range of automata that admit sound albeit expensive mathematical techniques, as opposed to focusing on a very narrow class of systems that often prematurely sacrifice genralizability for the sake of efficiency.

We built upon and integrated many existing ideas: e.g., semi-algebraic hybrid automata, the Blum-Shub-Smale model of "real" computation and TCTL (a powerful temporal logic formalism suitable for our setting)—more formally defined below.

Definition 1 Semi-Algebraic Set [45,47]. *Every quantifier-free boolean formula composed of polynomial equations and inequalities defines a semialgebraic set (i.e., unquantified first-order formulæ over the reals* - $(\mathbb{R}, +, \times, =, <)$*).* □

Definition 2 Semi-Algebraic Hybrid Automata [53]. *A k-dimensional* hybrid automaton *is a 7-tuple,* $H = (Z, V, E, Init, Inv, Flow, Jump)$*, consisting of the following components:*

- $Z = \{Z_1, \ldots, Z_k\}$ *a finite set of variables ranging over the reals* $\mathbb{R}$*;*
- (V, E) *is a directed graph of discrete states and transitions;*
- *Each vertex* $v \in V$ *is labeled by "Init"(initial), "Inv"(invariant) and "Flow" labels;*
- *Each edge* $e \in E$ *is labeled by a "Jump" condition;*
- *Init, Inv, Flow, and Jump are semi-algebraic.* □

Definition 3 Semantics of Hybrid Automata. *Let* $H = (Z, V, E, Init, Inv, Flow, Jump)$ *be a hybrid automaton of dimension k.*

- *A* location ℓ *of H is a pair* $\langle v, R\rangle$*, where* $v \in V$ *is a state and* $R \in \mathbb{R}^k$ *is an assignment of values to the variables of Z. A location* $\langle v, R\rangle$ *is said to be* admissible*, if* $Inv_v(R)$ *is satisfied.*
- *The* continuous reachability transition relation $\xrightarrow[\mathscr{C}]{h}$ *forces the state invariant to hold at every point except the end-point along the evolution curve determined by the flow equations during the* $h(>0)$ *time units from the current time* t_0*:*

$$\langle v,R\rangle \xrightarrow[\mathscr{C}]{h} \langle v,S\rangle \quad \textit{iff}$$
$$\Big(Flow_v(R,S,t_0,h) \wedge \forall S', h' \in [0,h)\ Flow_v(R,S',t_0,h') \Rightarrow Inv_v(S')\Big),$$

where $Flow_v(R,S,t,h)$ *is a relation between the continuous state R at time t and the continuous state S after h time units in the discrete state v. It is "well-defined" in the sense that* $\forall R,S,t,h\ Flow_v(R,S,t,h) \Rightarrow \{\forall h' \in [0,h)\ \exists S'\ Flow_v(R,S',t,h')\}$.

- *The* discrete reachability transition relation $\xrightarrow[\mathscr{D}]{0}$ *ensures that both parts of the* zero-time *jump*[1] *— the guard condition which needs to be satisfied just before the transition is taken, and the reset condition which determines the values after the transition, are satisfied.*

$$\langle v,R\rangle \xrightarrow[\mathscr{D}]{0} \langle u,S\rangle \quad \textit{iff} \quad \langle v,u\rangle \in E \wedge Jump_{v,u}(R,S).$$

- *The* transition relation $\mathscr{T}$ *of H connects the possible values of the system variables before and after* one step *— a discrete step for a time* $h = 0$ *or a continuous evolution for any time period* $h > 0$*:*

$$\mathscr{T}(\ell \xrightarrow{h} \ell') = \ \{h = 0 \wedge \ell \xrightarrow[\mathscr{D}]{0} \ell'\} \vee \{h > 0 \wedge \ell \xrightarrow[\mathscr{C}]{h} \ell'\}.$$

- *A* trace *of H is a sequence* $\ell_0, \ell_1, \ldots, \ell_n, \ldots$ *of admissible locations such that*

$$\forall i \geq 0,\ \exists h_i \geq 0,\ \mathscr{T}(\ell_i \xrightarrow{h_i} \ell_{i+1}).$$ □

Definition 4 Finite-Dimensional Machine Over $\mathscr{R}$: [13]. *A finite dimensional machine M over* $\mathscr{R}$ *consists of a finite directed connected graph with four types of nodes:* input, computation, branch *and* output.

In addition the machine has three spaces: input space $\mathscr{I}_M$, state space $\mathscr{S}_M$ *and* output space $\mathscr{O}_M$ *of the form* $\mathscr{R}^n, \mathscr{R}^m, \mathscr{R}^l$*, respectively, where* n, m *and* l *are positive integers.*

1. *Associated with the* input node *is a linear map* $I : \mathscr{I}_M \to \mathscr{S}_M$ *and a unique next node* β_1.
2. *Each* computation node η *has an associated* computation map*, a polynomial (or rational) map* $g_\eta : \mathscr{S}_M \to \mathscr{S}_M$ *given by m polynomials (or rational functions)* $g_j : \mathscr{R}^m \to \mathscr{R}, j = 1, \cdots, m$*, and a unique next node* β_η.

[1] $Jump_{v,u}(R,S) \equiv Guard_{v,u}(R) \wedge Reset_{v,u}(R,S)$.

3. *Each* branch node η *has an associated* branching function*, a nonzero polynomial function* $h_n : \mathscr{S}_M \rightarrow \mathscr{R}$.
4. *Each* output node η *has an associated linear map* $\mathscr{O}_\eta : \mathscr{S}_M \rightarrow \mathscr{O}_M$ *and no next node.* □

Theorem 1 Path Decomposition Theorem: [13]. *For any machine* $\mathscr{M}$ *over* $\mathscr{R}$ *the following properties hold.*

1. *For any* $T > 0$*, the* time-T halting set *of* $\mathscr{M}$*:* Ω_T $(= \bigcup_{\gamma \in \Gamma_T} \nu_\gamma)$ *is a finite disjoint union of basic semi-algebraic sets (respectively, basic quasi-algebraic sets, in the unordered case), where* Γ_T *is the* set of time-T halting paths *and* ν_γ *is the* initial path set.
2. *The* halting set *of* $\mathscr{M}$*:* Ω_M $(= \bigcup_{\gamma \in \Gamma_{M'}} \nu_\gamma)$ *is a countable disjoint union of basic semi-algebraic (respectively, basic quasi-algebraic) sets, where* $\Gamma_{M'}$ *is the* set of minimal halting paths.
3. *For* $\gamma \in \Gamma_M$ *(the* set of halting paths *of* $\mathscr{M}$*), the* input-output map Φ_M *restricted to* $\nu_\gamma - \Phi_{M|\nu_\gamma}$ *is a polynomial map, or a rational map if* $\mathscr{R}$ *is a field.* □

Definition 5 The Mandelbrot Set [44]. $\mathscr{M}$ *is the subset of the set of complex numbers* $\mathscr{C}$ *that remains bounded when subject to the following iterative procedure:* $f_0(c) = c$ *,* $f_{n+1}(c) = f_n(z)^2 + c$*. Formally, the complement* $\mathscr{M}'$ *of the Mandelbrot set is defined as*

$$\mathscr{M}' = \{c \in \mathscr{C} | f_n(c) \rightarrow \infty \text{ as } n \rightarrow \infty\}.$$

It is to be noted that $f_i(c) \geq 2$ *implies that eventually* $f_n(c) \rightarrow \infty$*.* □

Definition 6. The Mandelbrot Hybrid Automaton *consists of*

- *One discrete state with invariant* False *and two continuous variables x and y.*
- *Flow*$_1$: $\{ x' = x \wedge y' = y \}$ *(no continuous evolution).*
- *One Discrete State Transition:* $1 \rightarrow 1$ *with Jump*$_1$: $(x' = x^2 - y^2 + C_r) \wedge (y' = 2xy + C_i)$*, where* C_r *and* C_i *are two constants (real numbers).*
- *Only possible trace: zeno path of infinite self-loops.* □

Theorem 2 Undecidability Of The Mandelbrot Set: [13]. *The Mandelbrot set*[2] *cannot be expressed as the countable union of semi-algebraic sets over* $\mathscr{R}$*, and hence not decidable over* $\mathscr{R}$*.* □

Definition 7 TCTL[2]. *It has the following syntactic structure:*

$$\phi ::= p \mid \neg\phi \mid \phi_1 \vee \phi_2 \mid \phi_1 \exists\mathscr{U} \phi_2 \mid \phi_1 \forall\mathscr{U} \phi_2 \mid z.\phi.$$

Its associated semantics is described below:

- **z.**: *The freeze quantification "z." binds the associated variable z to the current time. Thus the formula* $z.\phi(z)$ *holds at time t iff* $\phi(t)$ *does.*

[2] The corresponding 2-dimensional set of real numbers.

– $\phi_1 \forall\mathscr{U}\, \phi_2$ *and* $\phi_1\ \exists\mathscr{U}\ \phi_2$*: universal (on all paths) and existential (on at least one path) "until" operators. For* $\phi_1\ \mathscr{U}\ \phi_2$ *to be true on a path,* ϕ_2 *is required to be true somewhere along the path, and* ϕ_1 *is required to be true all along the path up to (but not necessarily at) that point.* □

Remark 1. The basic notations are often extended by the following syntactic abbreviations [2].

1. $p\ \exists\mathscr{U}_{\leq max}\, q \equiv p\ \exists\mathscr{U}\ (q \wedge z.(z \leq max))$ and $p\ \forall\mathscr{U}_{\leq max}\, q \equiv p\ \forall\mathscr{U}\ (q \wedge z.(z \leq max))$: "subscripted" *Until* operators (*max* is the time-bound).
2. $(\forall\mathscr{F}\ p \equiv true\ \forall\mathscr{U}\ p)$ and $(\exists\mathscr{F}\ p \equiv true\ \exists\mathscr{U}\ p$: "eventuality" operators.
3. $(\forall\mathscr{G}\ p \equiv \neg\exists\mathscr{F}\neg p)$ and $(\exists\mathscr{G}\ p \equiv \neg\forall\mathscr{F}\neg p)$: "invariance" operators.

Definition 8 Single-Step Until Operator, $\triangleright$, [35]. *The formula* $p \triangleright q$ *holds if* $p \vee q$ *is true all along "one step" of the hybrid system and* q *is true at the end of the transition.* □

Definition 9 $T\mu$**-Calculus Syntax: [35].** $\phi ::= X \mid p \mid \neg\phi \mid \phi_1 \vee \phi_2 \mid \phi_1 \triangleright \phi_2 \mid z.\phi \mid \mu X.\phi$*, where* μ *is the* least-fixpoint *operator*[3]*. Thus,*

– Existential Until: $p\ \exists\mathscr{U}\ q\ =\ \mu X.(q \vee (p \triangleright X))$
– Universal Until:[4] $p\forall\mathscr{U}\, q\ =\ \neg(\neg q\ \exists\mathscr{U}\ (\neg p \wedge \neg q))$ □

3.1 What Questions Can and Cannot Be Answered

One may now wish to devise algorithmic algebraic solutions to various kinds of queries (in TCTL) to examine interesting properties and invariants about the hybrid automata that model biochemical systems. The simplest and perhaps the most important question that one can ask about these systems is the symbolic state reachability problem: namely, can one reach a particular state from an initial state by following the dynamics of the hybrid automaton which may be described symbolically? A more relevant biological question could be to provide a symbolic description of the initial conditions (states) from which the biological system (modeled via a semi-algebraic hybrid automaton) can reach a desired state (say, apoptosis state for a cancer cell), or avoid certain unsafe states. In this sense, algebraic descriptions in systems biology can be a potent tool. However, the immediate answers to these questions are depressingly negative. Thus, our community needs to engage in many years of focused work to devise a mature algebraic systems biological toolset. We and others have made some progress by exploiting approximations, bounded reachability analysis, etc. or by suitably constraining the power of the family of hybrid automata studied [54,52,50,17,15,49,51,48,14]. But much more remains to be done!

Just to summarize few of the positive steps in this direction, we mention the following two different approaches: The first way is to identify hybrid automaton classes for which the problem is decidable and to use such classes to model hybrid systems. In the last ten years, many decidable classes have been discovered [3,6,56,40,41,18], but, because of the restrictions imposed on them to achieve decidability, often they cannot be

[3] The *greatest-fixpoint* ν can be expressed as $\neg\mu X.(\neg\phi[X := \neg X])$.
[4] This translation is valid only when q is "finitely variable" over all premodels [35].

easily applied in the analysis of real biological systems. The second way is approximate analysis, like bounded model checking [31,27], abstract interpretation [4,5], or quotient reduction [32,33,34], to obtain a partial (or approximate) result for the model checking problem (e.g., the property holds for at least ten seconds starting from the initial condition).

On other approaches that resemble the systems described here, we enumerate few recent results: Anai [8] and Fränzle [26] independently suggested the use of quantifier elimination for the verification of polynomial hybrid systems. Anai and Weispfenning subsequently expounded the use of quantifier elimination for the reachability analysis of continuous systems with parametric inhomogeneous linear differential equations [9]. Fränzle went on to prove that progress, safety, state recurrence and reachability are semi-decidable using quantifier elimination of semi-algebraic formulæ [26], and to develop proof engines for bounded model checking [27]. Lafferiere et al. [42] have described a quantifier-elimination-centric method for symbolic reachability computation of linear vector fields. Many of these powerful techniques remain to be fully integrated into the context that systems biology proposes.

We only present technical details of the following negative result, here. Rest can be found in the reference [52].

Theorem 3 General Undecidability Of Reachability. *For semi-algebraic hybrid systems, reachability is undecidable even in Blum et al.'s "real" Turing machine formalism.*

Proof. Consider the Mandelbrot hybrid automaton defined earlier, with the complex number $C = C_r + \imath.C_i$. Let $S(t) = x(t) + \imath.y(t)$. After 1 discrete state transition (self-loop), we get

$$S'(t) = \{x(t)^2 - y(t)^2 + C_r\} + \imath.\{2x(t)y(t) + C_i\} = \{x(t) + \imath.y(t)\}^2 + \{C_r + \imath.C_i\}$$

In other words, $S'(t) = S^2(t) + C$ which is the defining equation of the *Mandelbrot Set*. Clearly, if there exists an evolution where $|S(t)| \geq 2$ then we know that C does not belong to the Mandelbrot set i.e. if the reachability query[5] $(x^2 + y^2 \geq 4)$ is decidable, it would imply that the Mandelbrot set is decidable, thus resulting in a contradiction. □

3.2 Final Thoughts

Lest some may mistakenly conclude that I have argued parochially in favor of theses over hypotheses (equivalently, Newton over Hooke), I conclude this review with the following beautiful quote from Hooke:

> "So many are the links, upon which the true Philosophy depends, of which, if any can be loose, or weak, the whole chain is in danger of being dissolved; it is to begin with the Hands and Eyes, and to proceed on through the Memory, to be continued by the Reason; nor is it to stop there, but to come about to the Hands and Eyes again, and so, by a continuall passage round from one Faculty to another, it is to be maintained in life and strength."

It is hoped that someday, algebra will serve its role as a strong link between biological theses and hypotheses— maintained in life and strength!

[5] $Reachable(p) \equiv \exists \mathscr{F}(p)$.

References

1. Alur, R., Belta, C., Kumar, V., Mintz, M., Pappas, G.J., Rubin, H., Schug, J.: Modeling and Analyzing Biomolecular Networks. Computing in Science and Engineering 4(1), 20–31 (2002)
2. Alur, R., Courcoubetis, C., Dill, D.: Model-Checking for Real-Time Systems. In: International Symposium on Logic in Computer Science, vol. 5, pp. 414–425. IEEE Computer Press, Los Alamitos (1990)
3. Alur, R., Courcoubetis, C., Halbwachs, N., Henzinger, T.A., Ho, P.-H., Nicollin, X., Olivero, A., Sifakis, J., Yovine, S.: The Algorithmic Analysis of Hybrid Systems. Theoretical Computer Science 138(1), 3–34 (1995)
4. Alur, R., Dang, T., Ivancic, F.: Reachability analysis of hybrid systems via predicate abstraction. In: Tomlin, C.J., Greenstreet, M.R. (eds.) HSCC 2002. LNCS, vol. 2289, pp. 25–27. Springer, Heidelberg (2002)
5. Alur, R., Dang, T., Ivancic, F.: Counter-example guided predicate abstraction of hybrid systems. In: Garavel, H., Hatcliff, J. (eds.) ETAPS 2003 and TACAS 2003. LNCS, vol. 2619, pp. 208–223. Springer, Heidelberg (2003)
6. Alur, R., Dill, D.L.: A theory of timed automata. Theoretical Computer Science 126(2), 183–235 (1994)
7. Amonlirdviman, K., Ghosh, R., Axelrod, J.D., Tomlin, C.J.: A Hybrid Systems Approach to Modeling and Analyzing Planar Cell Polarity. In: Proceedings of the 3rd International Conference on Systems Biology (2002)
8. Anai, H.: Algebraic Approach to Analysis of Discrete-Time Polynomial Systems. In: Proceedings of European Control Conference (ECC'99) (1999)
9. Anai, H., Weispfenning, V.: Reach set computations using real quantifier elimination. Technical Report MIP-0012, Fakultät für Mathematik und Informatik, Universität Passau (2000)
10. Antoniotti, M., Mishra, B., Piazza, C., Policriti, A., Simeoni, M.: Modelling Cellular Behavior with Hybrid Automata: Bisimulation and Collapsing. In: Priami, C. (ed.) CMSB 2003. LNCS, vol. 2602, pp. 57–74. Springer, Heidelberg (2003)
11. Antoniotti, M., Policriti, A., Ugel, N., Mishra, B.: XS-systems: eXtended S-Systems and Algebraic Differential Automata for Modeling Cellular Behavior. In: Proceedigs of the International Confernce on High Performance Computing, HiPC 2002, Bangalore, India (December 2002)
12. Batt, G., de Jong, H., Geiselmann, J., Page, M.: Qualitative Analysis of Genetic Regulatory Networks: A Model-Checking Approach. In: Bredeweg, B., Salles, P. (eds.) Working Notes of Seventeenth International Workshop on Qualitative Reasoning, QR-03, pp. 31–38 (2003)
13. Blum, L., Cucker, F., Shub, M., Smale, S.: Complexity and Real Computation. Springer, Heidelberg (1997)
14. Casagrande, A.: Hybrid Systems: A First-Order Approach to Verification and Approximation Techniques. PhD thesis, Department of Mathematics and Computer Science, University of Udine, Advisers - Policriti, A., Villa, T. (2006)
15. Casagrande, A., Mysore, V., Piazza, C., Mishra, B.: Independent Dynamics Hybrid Automata in System Biology. In: Proceedings of the First International Conference on Algebraic Biology, Tokyo (Japan), (November 28-30, 2005)
16. Casagrande, A., Mysore, V., Piazza, C., Mishra, B.: Independent Dynamics Hybrid Automata in Systems Biology. In: Proceedings of the First International Conference on Algebraic Biology (AB'05), November 2005, pp. 61–73, Tokyo, Japan, Universal Academy Press, Inc. (2005)

17. Casagrande, A., Piazza, C., Mishra, B.: Semi-Algebraic Constant Reset Hybrid Automata - SACoRe. In: Proceedings of the 44rd Conference on Decision and Control and European Control Conference (CDC-ECC'05), Seville, Spain, December 2005, pp. 678–683. IEEE Computer Society Press, Los Alamitos (2005)
18. Casagrande, A., Piazza, C., Mishra, B.: Semi-Algebraic Constant Reset Hybrid Automata - SACoRe. In: Proceedings of the 44rd Conference on Decision and Control and European Control Conference (CDC-ECC'05), Seville, Spain, December 2005, pp. 678–683. IEEE Computer Society Press, Los Alamitos (2005)
19. Chabrier, N., Chiaverini, M., Danos, V., Fages, F., Schächter, V.: Modeling and Querying Biochemical Interaction Networks. Theoretical Computer Science 325(1), 25–44 (2004)
20. Chabrier, N., Fages, F.: Symbolic Model Checking of Biochemical Networks. In: Proceedings of the First International Workshop on Computational Methods in Systems Biology, pp. 149–162 (2003)
21. Cornish-Bowden, A.: Fundamentals of Enzyme Kinetics, 3rd edn. Portland Press, London (2004)
22. de Jong, H.: Modeling and Simulation of Genetic Regulatory Systems: A Literature Review. Journal of Computational Biology 9(1), 69–105 (2002)
23. de Jong, H.: Modeling and Simulation of Genetic Regulatory Networks. Lectures Notes in Control and Information Sciences 294, 111–118 (2003)
24. Drulhe, S., Ferrari-Trecate, G., de Jong, H., Viari, A.: Reconstruction of Switching Thresholds in Piecewise-Affine Models of Genetic Regulatory Networks. In: Hespanha, J.P., Tiwari, A. (eds.) HSCC 2006. LNCS, vol. 3927, pp. 184–199. Springer, Heidelberg (2006)
25. Eker, S., Knapp, M., Laderoute, K., Lincoln, P., Meseguer, J., Sonmez, K.: Pathway Logic: Symbolic Analysis of Biological Signaling. In: Proceedings of the Pacific Symposium on Biocomputing, pp. 400–412 (January 2002)
26. Fränzle, M.: What will be eventually true of polynomial hybrid automata? In: Kobayashi, N., Pierce, B.C. (eds.) TACS 2001. LNCS, vol. 2215, pp. 340–359. Springer, Heidelberg (2001)
27. Fränzle, M., Herde, C.: Efficient proof engines for bounded model checking of hybrid systems. In: FMICS (2004)
28. Ghosh, R., Tiwari, A., Tomlin, C.: Automated Symbolic Reachability Analysis; with Application to Delta-Notch Signaling Automata. In: Maler, O., Pnueli, A. (eds.) HSCC 2003. LNCS, vol. 2623, pp. 233–248. Springer, Heidelberg (2003)
29. Ghosh, R., Tomlin, C.: Lateral Inhibition through Delta-Notch Signaling: A Piecewise Affine Hybrid Model. In: Di Benedetto, M.D., Sangiovanni-Vincentelli, A.L. (eds.) HSCC 2001. LNCS, vol. 2034, pp. 232–246. Springer, Heidelberg (2001)
30. Ghosh, R., Tomlin, C.: An Algorithm for Reachability Computations on Hybrid Automata Models of Protein Signaling Networks. In: Proceedings of the 44rd Conference on Decision and Control and European Control Conference (CDC-ECC'05), Seville, Spain, December 2005, pp. 2256–2261. IEEE Computer Society Press, Los Alamitos (2005)
31. Giorgetti, N., Pappas, G., Bemporad, A.: Bounded model checking of hybrid dynamical system. In: Proceedings of the 44rd Conference on Decision and Control and European Control Conference (CDC-ECC'05), Seville, Spain, December 2005, pp. 672–677. IEEE Computer Society Press, Los Alamitos (2005)
32. Girard, A., Pappas, G.J.: Approximate bisimulations for constrained linear systems. In: Proceedings of the 44rd Conference on Decision and Control and European Control Conference (CDC-ECC'05), Seville, Spain, December 2005, pp. 4700–4705. IEEE Computer Society Press, Los Alamitos (2005)
33. Girard, A., Pappas, G.J.: Approximate bisimulations for nonlinear dynamical systems. In: Proceedings of the 44rd Conference on Decision and Control and European Control Conference (CDC-ECC'05), Seville, Spain, December 2005, pp. 690–695. IEEE Computer Society Press, Los Alamitos (2005)

34. Girard, A., Pappas, G.J.: Approximate simulation relations for hybrid systems. In: Proceedings of Analysis and Design of Hybrid Systems (ADHA'06), June 2006, page Algero, Italy, June (to appear)
35. Henzinger, T.A., Nicollin, X., Sifakis, J., Yovine, S.: Symbolic Model Checking for Real-time Systems. In: 7th Annual IEEE Symposium on Logic in Computer Science IEEE, pp. 394–406. IEEE Computer Society Press, Los Alamitos (1992)
36. Hu, J., Wu, W.-C., Sastry, S.: Subtilin Production in Bacillus subtilis using Stochastic Hybrid Systems. In: Alur, R., Pappas, G.J. (eds.) HSCC 2004. LNCS, vol. 2993, pp. 417–431. Springer, Heidelberg (2004)
37. Ideker, T., Lauffenburger, D.: Building with a Scaffold: Emerging Strategies for High- to Low-Level Cellular Modeling. Trends in Biotechnology 21(6), 255–262 (2003)
38. Kappler, K., Edwards, R., Glass, L.: Dynamics in High Dimensional Model Gene Networks. Signal Processing 83, 789–798 (2002)
39. Keener, J.P., Sneyd, J.: Mathematical Physiology. Springer, New York (1998)
40. Kopke, P.W.: The Theory of Rectangular Hybrid Automata. PhD thesis, Faculty of the Graduate School, Cornell University, Adviser - Henzinger, T.A (1996)
41. Lafferriere, G., Pappas, G.J., Sastry, S.: O-minimal hybrid systems. Mathematics of Control, Signals, and Systems 13, 1–21 (2000)
42. Lafferriere, G., Pappas, G.J., Yovine, S.: Symbolic reachability computation for families of linear vector fields. Journal of Symbolic Computation 32(3), 231–253 (2001)
43. Lincoln, P., Tiwari, A.: Symbolic Systems Biology: Hybrid Modeling and Analysis of Biological Networks. In: Alur, R., Pappas, G.J. (eds.) HSCC 2004. LNCS, vol. 2993, pp. 660–672. Springer, Heidelberg (2004)
44. Mandelbrot, B.: The Fractal Geometry of Nature. Freeman Co, San Francisco (1982)
45. Mishra, B.: Algorithmic Algebra. Springer, New York (1993)
46. Mishra, B.: A Symbolic Approach to Modeling Cellular Behavior. In: Sahni, S.K., Prasanna, V.K., Shukla, U. (eds.) HiPC 2002. LNCS, vol. 2552, pp. 725–732. Springer, Heidelberg (2002)
47. Mishra, B.: Computational Real Algebraic Geometry, pp. 740–764. CRC Press, Boca Raton, FL (2004)
48. Mysore, V.: Algorithmic Algebraic Model Checking: Hybrid Automata & Systems Biology. PhD thesis, Department of Computer Science, New York University, Advisors - Mishra, B. (2006)
49. Mysore, V., Casagrande, A., Piazza, C., Mishra, B.: Tolque – A Tool for Algorithmic Algebraic Model Checking. In: The Ninth International Workshop on Hybrid Systems Computation & Control (HSCC06) Poster session (March 2006)
50. Mysore, V., Mishra, B.: Algorithmic Algebraic Model Checking III: Approximate Methods. In: Infinity 2005 – The 7th International Workshop on Verification of Infinite-State Systems, vol. 149(1) of Electronic Notes in Theoretical Computer Science, pp. 61–77 (February 2006)
51. Mysore, V., Mishra, B.: Algorithmic Algebraic Model Checking IV: Metabolic Networks. Journal of Mathematical Biology (to be submitted, 2006)
52. Mysore, V., Piazza, C., Mishra, B.: Algorithmic Algebraic Model Checking II: Decidability of Semi-Algebraic Model Checking and its Applications to Systems Biology. In: Peled, D.A., Tsay, Y.-K. (eds.) ATVA 2005. LNCS, vol. 3707, pp. 217–233. Springer, Heidelberg (2005)
53. Piazza, C., Antoniotti, M., Mysore, V., Policriti, A., Winkler, F., Mishra, B.: Algorithmic Algebraic Model Checking I: The Case of Biochemical Systems and their Reachability Analysis. CIMS-TR 2005-859, Courant Institute Of Mathematical Sciences (2005)
54. Piazza, C., Antoniotti, M., Mysore, V., Policriti, A., Winkler, F., Mishra, B.: Algorithmic Algebraic Model Checking I: The Case of Biochemical Systems and their Reachability Analysis. In: Etessami, K., Rajamani, S.K. (eds.) CAV 2005. LNCS, vol. 3576, pp. 5–19. Springer, Heidelberg (2005)

55. Platzer, U., Meinzer, H.-P.: Simulation of Genetic Networks in Multicellular Context. In: Kim, D.J., Polani, M.T (ed.) Fifth German workshop on Artificial Life: Abstracting and Synthesizing the Principles of Living Systems, pp. 43–51 Akad. Verl.-Ges. (2002)
56. Puri, A., Varaiya, P.: Decidebility of hybrid systems with rectangular differential inclusions. Computer Aided Verification, pp. 95–104 (1994)
57. Sachs, K., Gifford, D., Jaakkola, T., Sorger, P., Lauffenburger, D.A.: Bayesian Network Approach to Cell Signaling Pathway Modeling. In: Sci. STKE (2002)
58. Shmulevich, I., Gluhovsky, I., Hashimoto, R.F., Dougherty, E.R., Zhang, W.: Steady-State Analysis of Genetic Regulatory Networks Modelled by Probabilistic Boolean Networks. Comparative and Functional Genomics 4, 601–608 (2003)
59. Voit, E.O.: Computational Analysis of Biochemical Systems. A Pratical Guide for Biochemists and Molecular Biologists. Cambridge University Press, Cambridge (2000)
60. Voit, E.O., Savageau, M.: Equivalence between S-systems and Volterra systems. Mathematical Biosciences 78, 47–55 (1986)