HillTau: A fast, compact abstraction for model reduction in biochemical signaling networks.

Upinder S. Bhalla

NCBS-TIFR, Bellary Road, Bangalore, India

Abstract

Signaling networks mediate many aspects of cellular function, yet their complexity and experimental inaccessibility limit the completeness and precision with which we can specify them. The conventional, mechanistically motivated approach to modeling such networks is through mass-action chemistry, which maps directly to biological entities, facilitating experimental tests and predictions. However such models are complex, need many parameters, and are computationally costly. Here we introduce the HillTau form for signaling models. HillTau retains the direct mapping to biological observables, but in contrast it is sparse, uses far fewer parameters, and is extremely efficient to compute. In the HillTau formalism, the steady-state concentration of all reaction products is approximated by the Hill equation, and the dynamics by a time-course *tau*. We demonstrate its use in implementing several biochemical motifs, including association, inhibition, feedforward and feedback inhibition, bistability, oscillations, and a synaptic switch obeying the BCM rule. We show how HillTau models closely fit existing mass-action models 10 times their size, and run 100 times faster even with a Python implementation. The same data-fitting approach serves for model construction from experimental data. The major use-cases for HillTau are system abstraction, model reduction, scaffolds for data-driven optimization, and fast approximations to complex cellular signaling.

Keywords: Simulation, systems biology, mass-action, synaptic plasticity

Introduction

John von Neumann's elephant haunts mechanistically detailed models[1]. There are two major arguments to exorcise it: that mechanistic detail is needed to address certain kinds of questions; and that in the era of big data it is both easier and less biased to simply build up detailed models with all the available pieces. Here we describe a model formalism, the HillTau form, to help navigate between biological mechanisms and big data on the one hand, and the desirability of condensed model representations that expose the key principles of system function.

Cellular, and particularly synaptic signaling, is notoriously complex. There are an estimated 1400 protein species localized to the postsynaptic density alone (Bayés et al., 2011). These support a range of functions including synaptic transmission, maintenance, plasticity, activity-driven protein synthesis, metabolic control, and traffic (Bhalla, 2014b). Mass-action chemistry is a common denominator for mechanistically inspired modeling of these phenomena. This has the key virtue of defining specific biological entities (molecules) and processes (reactions) that map directly to experimental observables. Many studies (Bhalla and Iyengar, 1999; Shouval et al., 2002) are based at this level. Further levels of mechanistic detail include reaction-diffusion, stochastic chemistry mesoscopic stochastic methods with trapezoidal or cubic meshes (Wils and De Schutter, 2009; Oliveira et al., 2010) and even single-particle reaction-diffusion calculations (Stiles and Bartol, 2001; Andrews et al., 2010). Note that the additional mechanistic detail comes at a considerable computational cost.
A few studies have found ways to lessen the level of detail, typically by focussing on interactions without dynamics (e.g, (Sorokina et al., 2011)) or on dynamics with highly reduced interactions (e.g.,(Barak and Tsodyks, 2006)). Model detail may also be abstracted out through model reduction, which starts from a detailed (usually mass-action ODE) model and strips it down to core interactions needed to account for model behavior. There is a substantial body of work on model reduction techniques (reviewed in (Snowden et al., 2017)). Most of these methods retain the chemical kinetics formalism using ODEs to represent mass-action.
Biochemical signaling models frequently suffer from incomplete parameterization. Thus 'detailed' models of signaling pathways, which are of course essential for many kinds of mechanistic analyses and design of experiments, are often underconstrained. In this context, a reduced model is preferable as it requires fewer parameters. While some of the formal approaches to this yield very compact models, the mapping to experimental observables may be quite indirect (Snowden et al., 2017). Hence it is useful to have a compact chemically-inspired formulation to serve as the core for the model reduction while remaining easy to parameterize and predict using the same quantities that are measured in experiments (Maurya et al., 2005; Danø et al., 2006; Taylor et al., 2008). Indeed, a compact model with few parameters is arguably a better starting point to understand complex signaling with insufficient data, than is a mechanistically detailed model.

---

[1] John von Neumann was reported to have said: "With four parameters I can fit an elephant, and with five I can make him wiggle his trunk". This was a critique of overfitting models, and has been frequently cited by researchers of the abstract persuasion to support (over)simplification of models of complex biological systems, such as elephants (Mayer et al., 2010).

Savageau and colleagues have developed the Design Space Toolbox to facilitate a systematic approach to developing reduced signaling and transcriptional network models with specified properties such as multistability (Lomnitz and Savageau, 2016). They cast mechanistic models into a Generalized Mass Action form, and this is then analyzed to realize the required phenotypic repertoire. While this is an effective way of obtaining models with desired multi-state properties, it differs in objectives from our goal of having a reduced, very efficient representation of dynamic responses of complex reaction networks such as synaptic signaling.

Efficiency is a specific constraint in developing models of synaptic signaling. On the one hand, many neural functions depend on the nuances of signaling. For example, network properties are quite sensitive to different plasticity rules (Dan and Poo, 2004), neuromodulators (Roelfsema and Holtmaat, 2018), and mutations (Südhof, 2017). Network models also are expanding to include diffusible messengers controlling cellular activity and blood flow (Dormanns et al., 2016). At the single-cell level, explorations of receptor insertion and clustering (Hudmon et al., 2005; Wilson et al., 2016), sequence recognition (Bhalla, 2017) and synaptic tagging (Frey and Morris, 1997; Smolen et al., 2012) all require some level of reference to the chemical signaling. The crux of the problem arises when these studies need to scale beyond one synapse to whole-neuron (up to 10^4 synapses, (Bhalla, 2017)) or even network scales (e.g., 10^9 synapses (Markram et al., 2015)). Clearly, efficiency in memory and computations is important for such models.

The HillTau form addresses several key concerns with modeling of complex signaling networks. It utilizes only those observable states specified by the user to map directly to the chemistry, thus supporting sparse models that are easier to constrain with limited data. This requires very few parameters, yet behaves similarly to chemical cascades involving multiple intervening steps. Since the user specifies their chosen observables, each can be related directly to observations of concentration over time. The models are small and calculations are highly efficient, being closed-form and event-driven.

Methods and Results

HillTau formulation

The HillTau formulation assumes that the steady-state level Y_∞ of each signaling step (which may involve multiple chemical steps) is approximated by a Hill function

$$Y_\infty/Y_{input} = \Theta = L^n/(KA^n + L^n) \qquad \text{Eq 1.1}$$

Here Y_{input} is the input concentration to this signaling step, where Y_{input} is either predefined or coming from an upstream reaction. Likewise, the reactant L can either be predefined or come from an upstream reaction. We utilize a slightly modified form to permit the Ligand *L* to act in an inhibitory manner:

$Y_\infty/Y_{input} = 1-\Theta = 1 - L^n/(KA^n + L^n)$ Eq 1.2

We use a different equation to define steady-state behaviour of a system where a single substrate molecule Y_{input} is converted into a product Y:

$Y_\infty = Y_{input}^n / KA$ Eq 1.3

We use the equations 1.1-1.3 to approximate the steady-state behaviour of multiple reaction steps leading to the activation of molecule *Y* by stimulus *L*. Note that these equations are entirely feed-forward: the concentrations of substrate molecules L and Y_{input} are not affected by their participating in any downstream reactions.

Equations 1 define the steady state. We then assume that the approach of the system to steady-state is governed by a simple exponential with characteristic time τ (Fig 1A):

$Y/Y_\infty = 1 - \exp(-\Delta t / \tau)$ Eq 2

where Y/Y_∞ is the fraction of the steady-state level attained in time Δt. As a slight extension to this, we permit an optional separate time course τ_2 when Y is falling:

If $Y_\infty > Y(t)$ $Y/Y_\infty = 1 - \exp(-\Delta t / \tau)$ Eq 2.1
If $Y_\infty < Y(t)$ $Y/Y_\infty = 1 - \exp(-\Delta t / \tau_2)$ Eq 2.2

Putting Eq 1 and Eq 2 together, we have the following closed-form expression for the value of Y at time t + Δt:

$Y(t+\Delta t) = Y_{baseline} + Y(t) + (Y_\infty - Y(t)) * (1-\exp(-\Delta t/ \tau))$ Eq 3

The term $Y_{baseline}$ is an optional baseline level of molecule Y. Δt is the timestep. Note that this is a closed form: Δt can be as large as the end of the simulation. In the limit, of course,

$Y_{(t = \infty)} = Y_\infty + Y_{baseline}$ Eq 3.1

The required parameters for each reaction are *KA* and τ. $Y_{baseline}$ is optional and defaults to zero. τ_2 is optional and defaults to the same value as τ. The Hill Coefficient *n* and the flag for inhibition define the structure of the reaction system, hence they are not treated as parameters. Since equation 1 is governed by the Hill Equation, and Eq 2 describes the time-course τ, we call this chemical modeling framework the HillTau form.

A minimal-parameter version of the HillTau form need not use molecular concentrations, just fractional concentrations. In practice one typically uses molecular concentrations in standard experimental units such as μM (micromolar), since this permits direct mapping to the experimental observations for concentration and for KA.

For the rare cases where a non-chemical formulation is needed to describe the system, we provide an alternative algebraic expression for Y_∞:

$Y_\infty = f(Y1, Y2....)$ Eq 4

where f is an arbitrary algebraic function and Y1, Y2… are concentrations of other molecules. The use of this algebraic form is discouraged as it weakens the mapping of the model to the underlying chemistry.
After considerable experimentation, we made a design decision not to permit reactions with more than two species (Y_{input}, L) as reagents. In brief, the interpretation of the KA term became ambiguous, and the Hill formalism became confusing to apply. Multiple reagents can readily be accommodated using sequential reactions.

Computing time-evolution and steady-states.

To build complex reaction systems, we permit cascading of reactions so that any molecule can be a substrate or equation term in any other reaction. To reiterate, this is a purely feed-forward formulation, so substrates are not affected by any of their downstream targets. We obtain a dependency tree so that on each timestep the updates are carried out in an order which ensures that inputs ripple in order through the cascade of reactions. However, multi-step systems may include feedback. In such cases the program has to explicitly break the dependency chain, and then it becomes necessary for the modeler to specify a timestep Δt which is smaller than the time-course (τ) of the fastest reaction in the loop. Since one normally samples reactions at a time finer than the fastest reaction, this restriction usually has little impact on runtime. However, there are cases where the HillTau system is used to compute steady state values (e.g., in a dose-response curve). These could ideally be solved by taking an infinitely long time-step. Given the possible presence of feedback, we instead take a long, user-defined settling time and subdivide it into 10 equal steps so as to allow feedback reactions to also settle. A similar jump to user-defined time-points applies for event-driven calculations.

Model definition and reference implementation

HillTau reaction systems are set up in a simple JSON format (Fig 1C), for which we have provided a schema. We have implemented a small reference driver program in Python (hillTau.py) that loads the model, runs it with optional stimuli, and plots or saves the output. The hillTau.py file also provides a set of library functions for use in larger programs. The source file, schema file, examples and documentation are all available on GitHub (https://github.com/BhallaLab/HillTau). Python scripts for generating the figures in this paper are also provided.

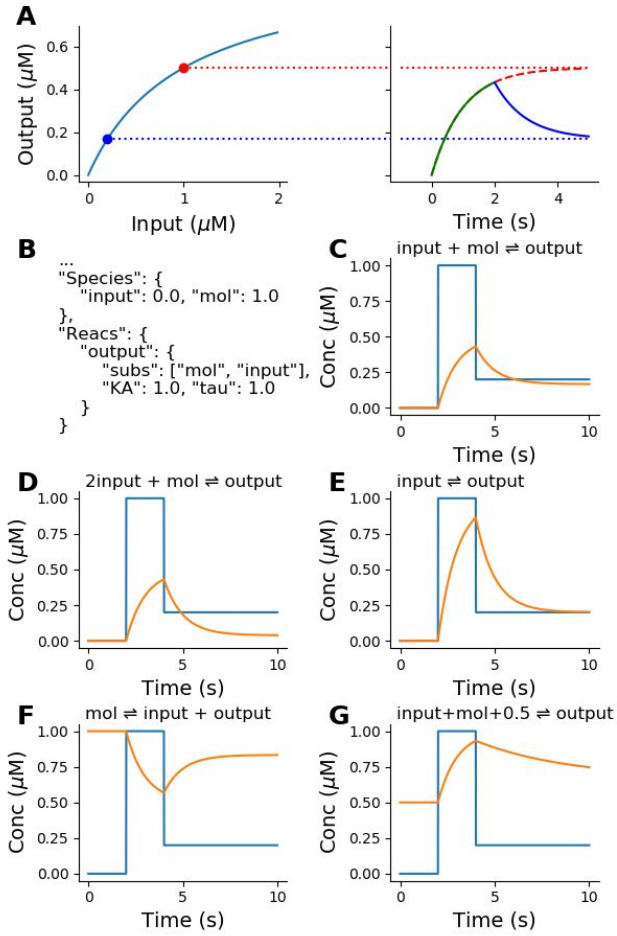

Figure 1. The HillTau formulation and representation of elementary chemical reactions.
A: Principle of HillTau formulation. Left: steady-state output values for different levels of the input molecule, computed by the Hill equation. In this and later simulations, two input values are used: first 1 μM (red dot) and later 0.2 μM (blue dot). Right: The simulator starts from the current value of the output, and computes the approach to the steady-state as an exponential time-course. Note that these are algebraic calculations, not numerical integration. In this example the output rises from zero toward the red dotted line for 2 seconds. Then the input is changed to 0.2 μM, and now the simulator approaches the steady-state value for this (blue dotted line) with an exponential time-course. B: JSON code snippet defining this reaction system. C-G: Inputs (blue) and simulated time-course of outputs (orange) for five different reactions. In all cases the HillTau output onset is identical to the output computed using numerical integration of a single reaction expressed as mass-action chemical kinetics. Decay time-course may differ from onset time-course in mass-action. C: Binding. D: 2nd order binding. E: Conversion. F: Inhibition, conceptually equivalent to removal of output molecules by binding of input to the output molecule, and sequestration of the resultant complex. G: Variant of binding reaction, in which there is a fixed baseline of 0.5 μM, and the system has different on (tau = 1s) and off (tau2 = 5s) time courses. This is a common outcome of competitive on and off processes.

The HillTau form can model a range of chemical signaling motifs

We implemented a range of elementary chemical signaling functions to illustrate the use of the HillTau form (Figure 1). The HillTau versions of most of these reactions have an exact fit to their mass-action counterparts (Supplementary Figure S1.1). We further implement key signaling motifs, including feedback inhibition, oscillation, and bistables (Figure 2). In the case of the first two we compared the HillTau approximation with mass-action versions, where the fit is not perfect, but the models are substantially reduced.

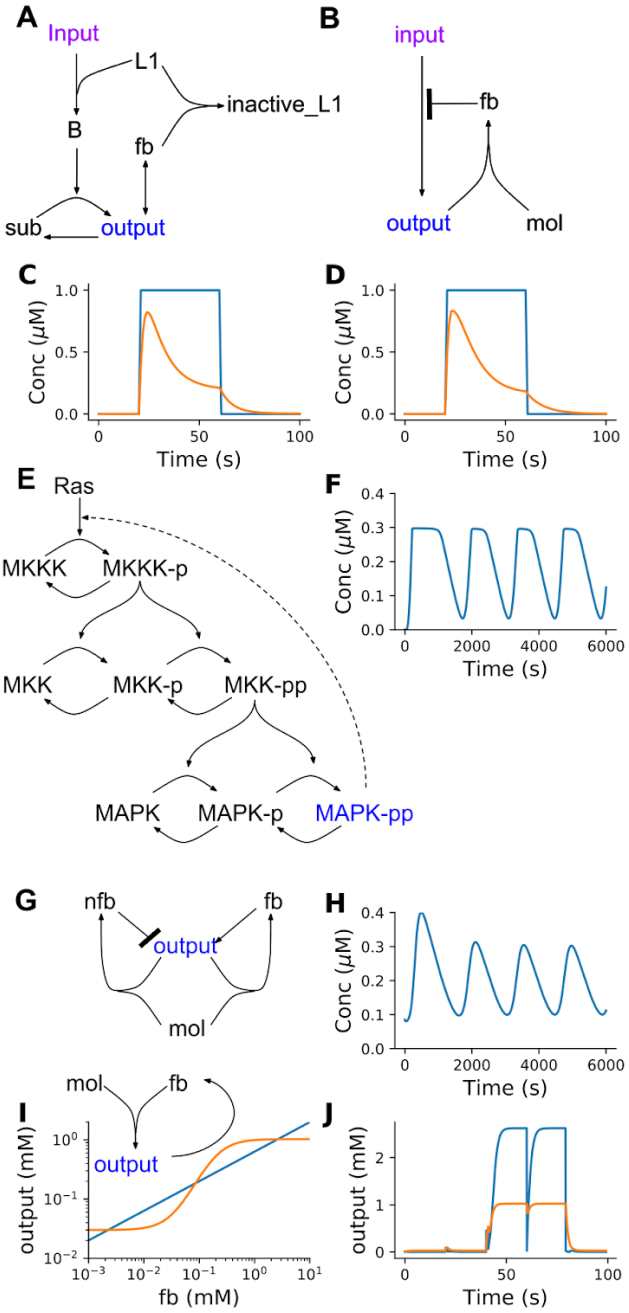

Figure 2. HillTau models of key signaling motifs. In all reaction schemes the molecule in blue is the readout used in the graphs. A-D: Feedback inhibition. Blue curve is stimulus, orange is response. A: Mass-action reaction scheme for feedback inhibition, involving 7 molecules and 5 reactions. B: HillTau version, using 4 molecules and 2 reactions. C, D: Simulations for mass-action (C) and HillTau versions of feedback inhibition. E-G: Oscillator. C: Mass-action model based on ultrasensitive MAPK cascade, taken from (Kholodenko, 2000). This uses 15 molecules, and 11 reactions. F: Simulation of mass-action oscillator. G: HillTau scheme, using 4 molecules and 3 reactions. H: Simulation of HillTau oscillator. I,: Bistability. I: Null-cline intersection plot showing stable states of system as the intersection points. This was generated by varying the feedback molecule fb, and measuring output (orange), and then varying output and measuring fb (blue). Inset: HillTau reaction scheme. J: Time-series illustration of state switching in the bistable. It starts in the low state, at 20 s a small excitatory input is given which fails to switch the state. At 40 s a strong input causes switching to the high state. At 60 s a weak inhibitory input fails to turn it off, but at 80 s a strong

inhibitory input returns the state to baseline. Excitatory and inhibitory inputs were delivered by transiently setting the level of fb to high or low values.

We then explored the use of HillTau for two synapse-related models, the BCM curve, and a synaptic switch triggered by a mechanism similar to the BCM curve, yielding bidirectional long-term state changes triggered by calcium (Figure 3). In later sections we examine HillTau versions of large mass-action models of activity-driven protein synthesis, and explore the potential for model reduction.

The HillTau form compactly represents bidirectional synaptic plasticity

Synaptic plasticity is one of the most-modelled neuronal signaling processes (Lisman, 1989; Bhalla and Iyengar, 1999). The key features that have been represented include stimulus strength-dependence, timing dependence, and long-term state storage (Bhalla, 2014a). A few studies have come up with rather detailed models to implement each of these processes (Hayer and Bhalla, 2005; Smolen et al., 2012; Kim et al., 2013). As an illustration of all these properties in the HillTau system, we implemented bidirectional synaptic plasticity including long-term synaptic state changes (Figure 3). One of the interesting aspects of synaptic plasticity is that in many systems, the same input modality (typically read out as Ca^{2+} concentration) can give rise to both synaptic depression and potentiation. This has significant theoretical implications and an abstract rule for this bidirectional plasticity was proposed by Bienenstock et al (BCM rule,(Bienenstock et al., 1982)). We first devised an already simplified mass-action version of the BCM rule using 9 molecules and 6 reactions (Figure 3A). Here, resting Ca^{2+} does not alter the state of the model; low Ca^{2+} causes depotentiation (that is, reduction of receptor levels), and high Ca^{2+} causes potentiation (Figure 3C-F). We then fit both the transient and the steady-state properties of the mass-action using just 3 reactions in HillTau (Figure 3B, 3C-F). As a further elaboration, we next introduced a bistable switch for long-term retention of synapse state, which was driven bidirectionally by the BCM rule (Figure 3F). The bistable switch, derived from CaMKII signaling, controls receptor insertion (Figure 3 E, F). Using this model we delivered a typical LTP stimuli (strong but brief Ca^{2+} input), leading to sustained synaptic AMPAR elevation. We followed this with a typical LTD stimulus (modest but sustained Ca^{2+} input), which turned the switch off again and led to reduction in AMPAR. This composite model required 4 reactions and one equation in the HillTau form, the equation handling the CaMKII feedback that gives rise to bistability. Several mass-action models of synaptic state switches include these elements (e.g., (Lisman, 1989; Bhalla and Iyengar, 1999; Hayer and Bhalla, 2005; Singh and Bhalla, 2018)) and they typically involve far more molecules and reactions (e.g., the Hayer and Bhalla 2005 model used 133 molecules and 215 reactions).
Can we create a reverse mapping from these simplified HillTau models to mass-action chemistry? As a proof-of-principle, we demonstrate a reverse mapping from HillTau to mass-action for the BCM curve simulations in Fig 3 C, D, E. This is described in the supplementary material S3.1 and in supplementary Figure S3.2. This mapping results in extra reactions and numerically stiff ODE systems, but illustrates how the reverse chemical mapping could take place.

Overall, these examples illustrate how compact HillTau models can represent both the bidirectional induction of plasticity, and also long-term maintenance of synaptic state.

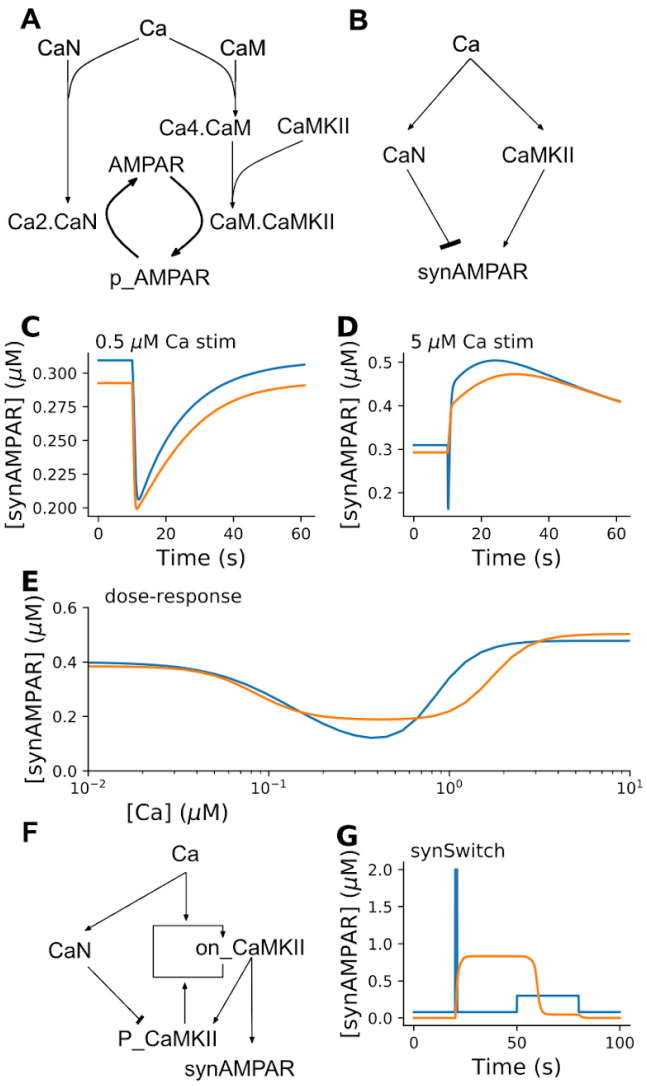

Figure 3: HillTau version of synaptic plasticity rules. A. Mass-action model for generating Beinenstock-Cooper-Munro (BCM) rule for synaptic plasticity. Calcium triggers both an inhibitor (CaN) and a stimulus (CaMKII) for receptor phosphorylation and insertion into the synapse. CaN activates at lower [Ca^{2+}], so there is initially a reduction in p_AMPAR. CaMKII is present at very high levels, so at higher Ca^{2+} it outcompetes CaN to give an increase in p_AMPAR. B. BCM rule implemented in HillTau. C-E: Comparison of mass-action model with HillTau model. Orange is HillTau, blue is mass action. C: 1s stimulus at 0.5 μM Ca^{2+} gives a reduction in synAMPAR. D: 1s stimulus at 5 μM Ca^{2+} gives an increase in synAMPAR. E: Full dose-response curve of steady-state synAMPAR as a function of [Ca^{2+}]. F: Schematic for BCM rule model feeding into bistable model, implemented in HillTau. G: time-course of simulation of bidirectional plasticity using different Ca^{2+} stimuli. At t =20, a 1s stimulus of 2μM Ca^{2+} causes a transition to the active state. At t = 50s, a 30s stimulus of 0.3 μM Ca^{2+} pulls the system back to resting state.

HT models can be optimized to fit biochemical measurements

The above examples illustrate how HillTau can represent biological signaling motifs, and build them up into networks with interesting computational properties. We next approached a complementary problem in signaling, to take a complex signaling system, and fit simple HillTau-form models to it as a way to infer computational properties. The basic flowchart is illustrated in Figure 4. This approach works in the same way for model construction from experimental response curves, and for model reduction from response curves taken from detailed models.

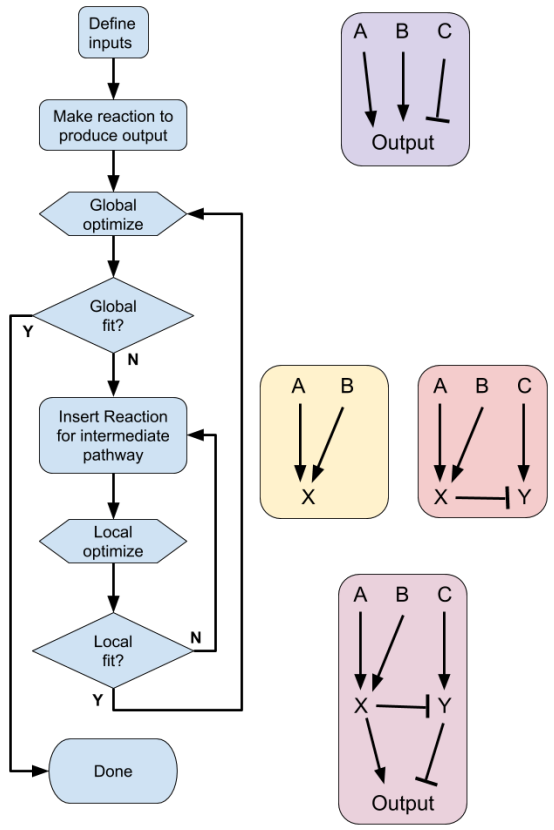

Figure 4: Flowchart for model building using HillTau. Left: flowchart. Right top: initial reaction with inputs A, B, and C. Right middle: successive local increments to model, introducing reactions X and Y respectively. Right lower: Final HillTau scheme with good fit at all stages.

As an example of this flowchart and the use of HillTau fitting to match direct biochemical data, we derived a HillTau model of synaptic activity-triggered protein synthesis. Our reference data was obtained by running a series of 'experiments' on a published model implemented in mass-action kinetics (Jain and Bhalla, 2009). The original model was based on numerous experiments, and included 123 molecules and 120 reactions (Fig 5A). The input pathways were Ca^{2+} and BDNF, and the final output was protein synthesis rate.

We obtained six ‘experimental’ curves including both time-series and steady-state experiments (Figure 5D-I). We utilized the FindSim framework
(Viswan et al., 2018) to run the HillTau model, to compare its results with each of these ‘experiments’, and to generate a score to measure goodness of fit. This score was used as the objective function in a gradient-descent optimization method from Python (scipy.optimize module, method “L-BFGS-B”,(Zhu et al., 1997)). Although our dataset was obtained from an existing model, exactly the same experiment-fitting approach could be used for direct experimental data.
As a first pass (top of the flowchart in Fig 4), we simply implemented the two input pathways connecting to the readout pathway (Fig 5B). This generated responses which were qualitatively similar to the reference data, but not very accurate. (Supplementary Figure S5.1). We then repeated the same set of simulated ‘experiments’ and selected two intermediate pathways as readouts, based on their known distinct roles in protein synthesis signaling. The selected pathways were S6K and CaMKIII, which respond primarily to BDNF and Ca^{2+} respectively, though of course there is cross-talk. We optimized S6K and CaMKIII to the inputs (Supplementary Figures S5.2 and S5.3) and obtained reasonable fits. These steps correspond to the middle of the flowchart in Figure 4. Finally, we made a composite model incorporating the inputs, S6K, CaMKIII and the protein readout (Figure 5C), and having 7 molecules and 8 reactions. This fit well for most experiments (Figure 5D-I), with the exception of the response to rapid Ca^{2+} pulses, where the HillTau model low-pass filtered the fast transient responses (Figure 5F).

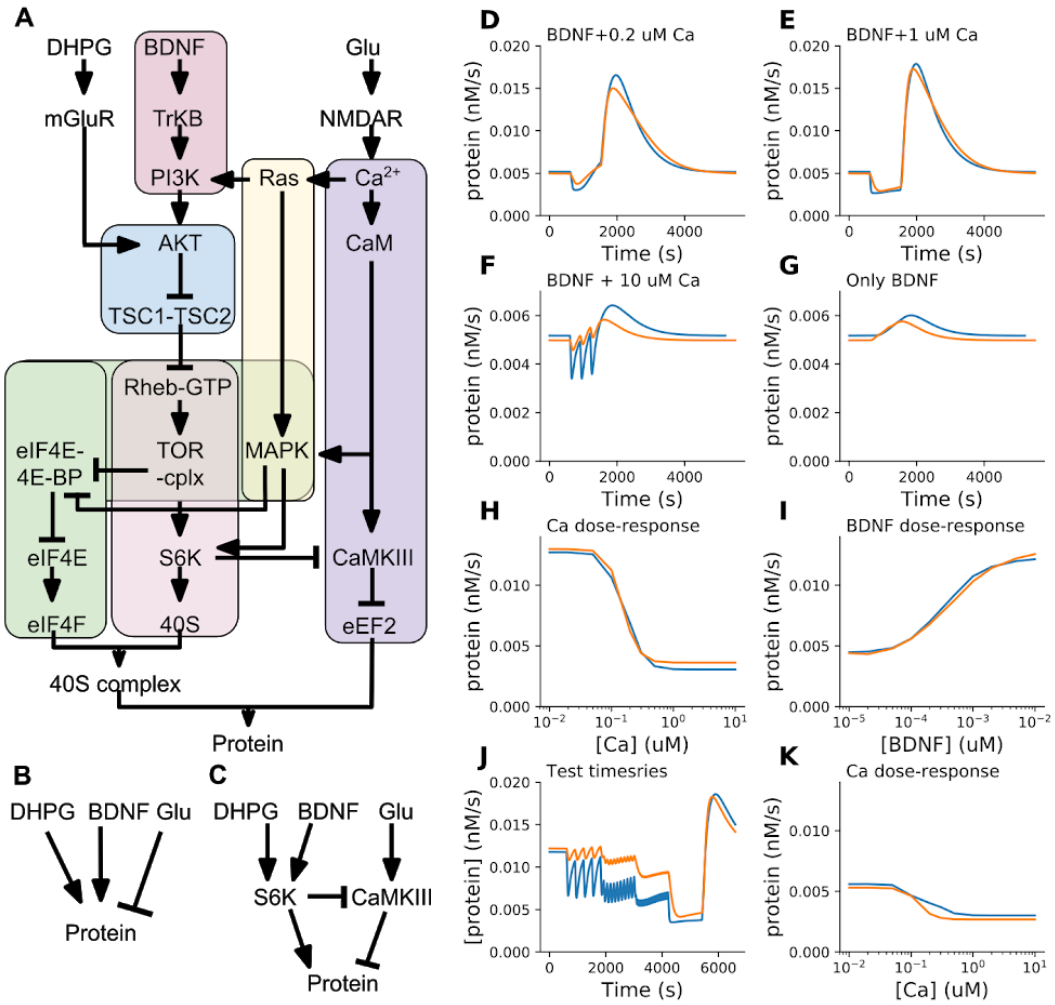

Figure 5: Model fitting and model reduction. A: Block diagram of source model with 123 molecules and 120 reactions, from (Jain and Bhalla, 2009). B: First pass reduced model in HillTau. C: Eventual reduced HillTau model. D-I: Six 'experiments' on reference model, used to tune parameters for HillTau version. In all cases protein production rate is readout. Blue plots are reference, orange are HillTau. D: BDNF@3.7 nM + Ca^{2+}@0.2 μM, 900 seconds. E: BDNF@3.7nM, Ca^{2+}@1μM. F: 3 pulses of BDNF@3.7 nM for 5s, coincident with Ca^{2+}@10μM for 1s, pulses separated by 300 s. G: Same as F, but Ca^{2+} held at baseline of 0.08 μM. H: Dose-response of protein vs. Ca^{2+}, holding BDNF fixed at 3.7 nM. I: Dose-response of protein vs BDNF, holding Ca^{2+} fixed at 0.08 μM. J, K: Two test simulations to see how the reduced HillTau model generalizes. J: Protein production rate for fixed BDNF@3.7 nM, where Ca^{2+} was given in 1 second pulses at intervals of 300, 120, 60 and 10 seconds; each pulse sequence lasting for 1200 s. The last stimulus and the rebound fit accurately, but there was only a qualitative fit to the lower-frequency pulses. K: Dose response of protein production rate vs. Ca^{2+}, holding BDNF at basal levels of 0.05 nM. The fit was reasonably good.

As a test of the ability of the model to generalize, we ran a series of test 'experiments' which had not been used for fitting the HillTau model. First, we held BDNF constant at 3.7 nM and delivered a series of 10 μM Ca^{2+} inputs as 1s pulses at different intervals (Fig 5J). As anticipated from Fig 5E, the HillTau transients were smaller than in the original model. The mean response was qualitatively similar at long intervals, and achieved a close quantitative match for the high-frequency Ca^{2+} input and subsequent recovery transient

(Figure 5J). Second, we did a dose-response curve where we held BDNF at resting levels (0.05 nM) and varied [Ca^{2+}] (Fig 5K). Here the fit was quite close.
In summary, we developed a flowchart for developing HillTau models to fit complex responses of experimental and simulated signaling pathways. Using this flowchart, we iteratively developed a HillTau model of 7 molecules and 8 reactions to fit a mass-action model having 123 molecules and 120 reactions.

HillTau models are compact and efficient

We next took a set of HillTau models of various levels of complexity, and compared various measures of computational cost with the mass-action equivalents (Table 1, Figure 6.)

We first compared model complexity, measured as the number of parameters needed to specify the model. The number of parameters scales roughly as

*# of molecules + 2 * # of reactions*.

This is a slight overestimate, since some of the molecules are state variables and we do not need concentration values for them. Each reaction needs two parameters, Kf and Kb for conversion reactions, and Km and kcat for enzymes. We sampled from among the mass-action models presented in the above sections, having 3 to over 360 parameters, and included an additional model with almost 750 parameters. (Fig 6A). The HillTau form had a similar scaling with molecules and reactions, except that HillTau also allows for an optional *baseline* term in the reactions so the average scaling is slightly larger than *2 * # of reactions*. We found that the HillTau form became increasingly effective at model reduction for larger models. Note that here the optimization goal was to obtain a single end-point response. Further reactions would be needed to also represent intermediate pathway readouts.

We then examined run-time efficiency. As a simple and unfair comparison, we took run-times for the C++ compiled and optimized code for solving the ODE systems of mass-action chemistry in MOOSE, and compared it with the Python implementation of matching HillTau models. For small models, the ODE calculations were about twice as fast as HillTau, but for large models HillTau became over 500 times faster (Figure 6B). This was sensitive to numerical issues such as stiffness of the ODEs describing the mass-action systems. Interestingly, the runtime for HillTau models grew almost linearly with the number of parameters (Fig 6C, slope = 0.86, R^2 = 0.93). This is because the HillTau calculations are not sensitive to numerical stiffness, in contrast to variable-timestep methods for ODE solutions used for mass-action calculations.

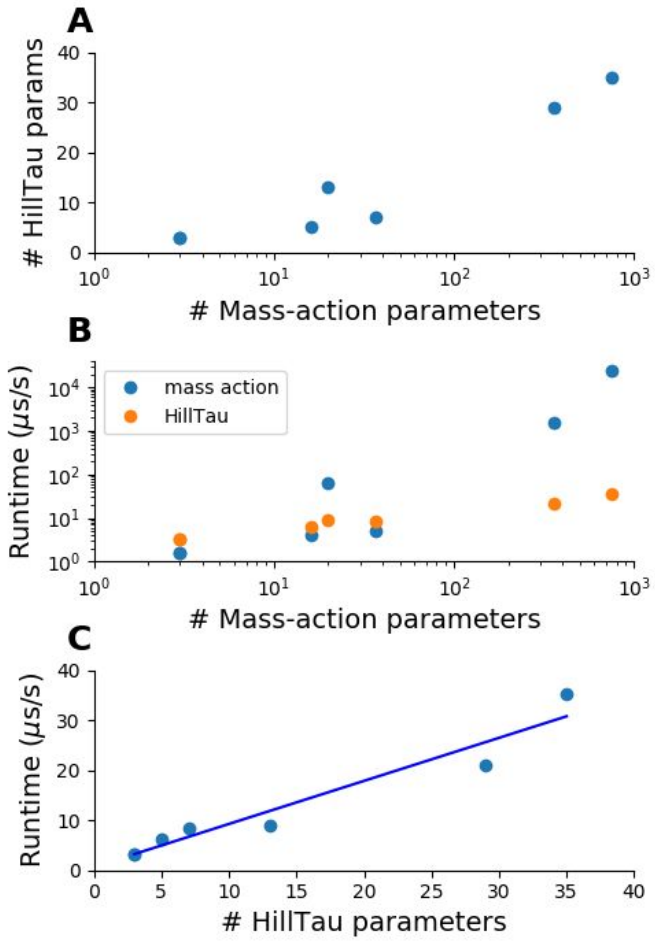

Figure 6. Efficiency of HillTau method. A. Scaling of number of parameters. The HillTau form uses far fewer parameters than mass-action, and becomes more efficient with larger models. B, C: runtime scaling. All runtimes are expressed in μs of wallclock time to execute 1 second of simulation time. B. Comparison of runtimes of the HillTau and mass-action forms, where a Python script was used to compute HillTau and C++ (MOOSE) for mass action. Above about 100 parameters, the HillTau form has a huge efficiency advantage. C. HillTau model runtime scales linearly with number of parameters. R^2 = 0.93, slope = 0.86, intercept = 0.66, units in μs/s.

Dose-response experiments are particularly efficient to compute using HillTau. The conventional way to do these for mass-action is to run an ODE system out to steady-state for each successive dose. This may take a long while especially if the system is stiff or converges slowly. It is also possible to use linear algebraic root-finding to find the steady-state value in one step, but this is complex to set up and may encounter numerical problems such as non-invertible matrices (Clarke, Bruce L., 1981). In HillTau, the form itself incorporates the steady-state value, so in principle one could leap to the steady value in one step. To be more conservative, the HillTau does so in 10 steps to smooth out transients and to allow any feedback signals to propagate through the system.

Numerically, the speedups are about 100 times greater than those in Figure 6B. This is because typical settling times use 1000 s simulation time for ODEs vs 10 steps for HillTau. Thus HillTau was between 1 and 4 orders of magnitude faster than ODE numerical integration for this settling calculation, and the speed up grew larger for larger models.

Overall, HillTau models are compact and highly efficient compared to ODE-solved mass-action models. The efficiency improves for larger models.

Discussion

We have designed HillTau, a compact, computationally efficient abstraction of chemical signaling that is effective in building models that sample a few nodes in a complex signaling network. It uses an event-driven algebraic representation based on the Hill equation and exponential relaxation to steady state. HillTau is effective in representing a range of chemical signaling motifs and complex synaptic models, using biological observables of molecules, reactions, association constants and time-courses. We show its applicability for model reduction by optimizing the fit of its responses to those of a reference mass-action model. This generates very compact models. The identical optimization approach works to build a HillTau model directly from experimental data. Thus HillTau addresses many of the concerns of model-building with limited data, and serves as a scaffold for eventual development of more detailed models.

Capabilities and limitations

HillTau is phenomenological and semi-heuristic, in that it uses the Hill equation to achieve concentration dependencies that fit well, but ignores many intervening chemical steps. This combination gives it the strong points indicated above, namely speed, compactness, and consistent mapping to experimental observables, but it also sets out clear limitations. Foremost among these is that it can only make limited mechanistic predictions, since it is missing many mechanistic steps. For example, a HillTau model would be limited in its ability to predict drug targets or side-effects because it may have lumped together potential molecular targets into a single reaction step. It is, however, quite effective in representing and predicting emergent signaling properties because it captures dynamics and topology of signaling networks.

The current HillTau formulation is limited in its handling of two important aspects of signaling in neurons: stochasticity and diffusion. These phenomena are out of the scope of our current implementation, which has focussed on development, validation, simplicity and speed. Many biochemical signaling processes experience substantial stochasticity, particularly in small-volume systems such as the synapse which is a target of our modeling. One possible way to introduce stochasticity would be through the linear noise approximation of the chemical Langevin equation (Wallace et al., 2012), which if used in an event-driven manner could be quite efficient. We anticipate it will take extensive validation to establish its utility in the HillTau framework. Similarly, there are potential ways to elaborate upon HillTau to use an event-driven approximation to diffusion, but these will require later follow-up.

Based on these attributes, we discuss four major use-cases for HillTau: model reduction, system abstraction, scaffolds for data-driven optimization, and efficient approximations to complex cellular signaling.

Model reduction

The HillTau form provides the capability to systematically scale down complex signaling networks. Several algorithmic approaches have been brought to bear on this problem, including collapsing multiple mass-action steps into one (Radulescu et al., 2008), and power-law generalizations of mass-action signaling (Savageau, 2001). With HillTau one can use a well-known heuristic/optimization approach to simplifying large networks ((Quaiser et al., 2011; Apri et al., 2014)). The approach reported in our current study relies on the modeler starting from a minimal input-> output mapping, then iteratively picking relevant major nodes in the network, and optimizing each time to fit the data. The actual optimization calculations can be done quickly due to HillTau's efficiency. Thus one can converge on the minimal set of intermediate nodes (illustrated in Figures 4 and 5, and supplementary figures S5.1, S52. and S5.3) to achieve the desired accuracy of model fit to data. Like other model reduction approaches, this minimal set of nodes is a good compromise between available data and model accuracy (Snowden et al., 2017). Unlike several other reduction approaches, HillTau retains a direct mapping to observable biological entities. Indeed, the HillTau representation of a signaling node is closer to the conventional intuition based on pathway schematics, than is a full mass-action reaction representation. Like pathway diagrams, each HillTau reaction receives excitatory or inhibitory inputs, and may receive 'dashed line' inputs representing several intermediate steps. Further, pathway diagrams typically assume implicit back reactions and decay of activity when stimuli are removed. This too is built into how HillTau reactions work. In contrast to these simple mappings from pathway diagrams to the HillTau form, it is often difficult to map between signaling diagrams and the full mass-action reaction schemes (Bhalla, 2002, 2003). While previous model reduction studies have worked on different pathways, the survey of methods in (Snowden et al., 2017), suggests that HillTau achieves as good or better model reduction for large models than most other methods.

System abstraction and functional modules

Putting model reduction together with its effective mapping to abstract pathway diagrams, we propose that HillTau forms a useful tool for arriving at functional modules in complex signaling networks. Such modules have long been considered a conceptual basis for understanding complex signaling ((Bhalla and Iyengar, 2001)). Typically they have been ascertained by manual inspection and dynamical analysis of components of signaling networks, for example, the nested feedback loops in the cell cycle (Novák and Tyson, 2004). A more scalable approach to uncovering such modules has been to use graph theory for motif analysis on detailed mass-action models, but this approach loses key aspects of system dynamics (Alon, 2007). With the HillTau formalism in conjunction with the flowcharts for model reduction, we are able to generate highly reduced reaction graphs that nevertheless support rather accurate dynamics. Thus HillTau supports a data-driven approach to arrive at functional modules.

While functional modules are good for analysis, we note that biology does not necessarily partition signaling networks into neat modules (Bhalla, 2003). Indeed, cross-talk between pathways is common. Nevertheless, the clean feed-forward signaling in HillTau is possible in principle. As a proof-of-principle, we demonstrate a reverse mapping from HillTau to mass-action, though this results in larger and numerically stiff ODE systems (Supplementary material and supplementary Figure S3.1). The key aspect of this reverse mapping is the requirement for output 'buffers' after each reaction step, which keeps the signal flow unidirectional. The buffers are implemented as fast futile enzyme cycles

catalyzed by the output molecule. This shows up an interesting contrast between engineered (HillTau) and biologically inspired (mass-action) formalisms to do the same signaling tasks (e.g, Figures 2, 3, 5). The HillTau systems are smaller and their unidirectional information flow makes them simpler. The equivalent biological systems don't select for simplicity (as imagined by an engineer). Further, biological signaling rarely utilizes the metabolically expensive futile reactions that would be needed for unidirectional information flow.

Scaffolds for data-driven optimization of ODE models

Model fitting of large mass-action models is difficult in at least two ways: there are typically far fewer experiments than parameters, and it is computationally costly to run a large ODE model many times for carrying out an optimization approach to model fitting. We propose that the HillTau form may provide a useful bridge on both these counts. As we have illustrated above, HillTau models lend themselves to fitting to experiments because they have few parameters and they run quickly. Several advantages accrue from an initial pass to make and fit a HillTau model. 1. In building and optimizing a HillTau model, the optimization dataset will be use-tested, and gaps identified. 2. The essential pathway structure of the model will be defined by the HillTau model, and key interactions identified. The mass-action model must, at minimum, incorporate these interactions. 3. The parameters of the HillTau model set bounds for those of the detailed reaction sets. For example, the time-course of any individual mass-action reaction step must be faster than the HillTau reaction in which it is embedded.

Computationally efficient approximations to cellular signaling

One of the key use-cases envisioned for HillTau was to model complex cellular signaling, with synaptic signaling as an exemplar. Several of the examples in the current paper are in this domain, specifically the BCM curve and the coupled BCM curve with bistables (Figure 3), and the synaptic protein synthesis pathway (Figure 5). While even these large ODE signaling models run somewhat faster than wall-clock time (Figure 6), there are at least two cases where far greater efficiency is desirable. First, as mentioned above, model parameter optimization requires a large number (100s to 1000s in our experience) of evaluations of complete simulations. Each of these synaptic simulations may have to run for many thousands of seconds of simulation time to compare with typical plasticity experiments (Ajay and Bhalla, 2004). Further, one typically optimizes a pathway to fit numerous experiments, all of which must be simulated for each evaluation. Together, this is computationally expensive.

Second, a single neuron may have over 10,000 spines, and there may be many such neurons in a network. If each synapse is to implement a complex biochemical pathway the computational costs may be formidable. Network plasticity models (Higgins et al., 2014), and cellular sequence selectivity models (Bhalla, 2017) are examples of this scale of model. Indeed, (Higgins et al., 2014) have used an efficient event-driven calculation of synaptic weights with a similar exponential decay calculation as in HillTau. HillTau signaling provides a way to implement far more complex synaptic dynamics in every synapse, even in large networks. With compiled implementations of the HillTau form, this should incur very small additional computational costs over the regular electrical calculations.

Acknowledgments

USB received support from NCBS-TIFR and the Department of Atomic Energy, Government of India, under project no. 12-R&D-TFR-5.04-0800. USB is a JC Bose Fellow.

References

Ajay, S. M., and Bhalla, U. S. (2004). A role for ERKII in synaptic pattern selectivity on the time-scale of minutes. *Eur. J. Neurosci.* 20, 2671–2680. doi:10.1111/j.1460-9568.2004.03725.x.

Alon, U. (2007). Network motifs: theory and experimental approaches. *Nat. Rev. Genet.* 8, 450–461. doi:10.1038/nrg2102.

Andrews, S. S., Addy, N. J., Brent, R., and Arkin, A. P. (2010). Detailed simulations of cell biology with Smoldyn 2.1. *PLoS Comput. Biol.* 6, e1000705. doi:10.1371/journal.pcbi.1000705.

Apri, M., Gee, M. de, Mourik, S. van, and Molenaar, J. (2014). Identifying Optimal Models to Represent Biochemical Systems. *PLOS ONE* 9, e83664. doi:10.1371/journal.pone.0083664.

Barak, O., and Tsodyks, M. (2006). Recognition by variance: learning rules for spatiotemporal patterns. *Neural Comput.* 18, 2343–2358. doi:10.1162/neco.2006.18.10.2343.

Bayés, A., van de Lagemaat, L. N., Collins, M. O., Croning, M. D. R., Whittle, I. R., Choudhary, J. S., et al. (2011). Characterization of the proteome, diseases and evolution of the human postsynaptic density. *Nat. Neurosci.* 14, 19–21. doi:10.1038/nn.2719.

Bhalla, U. S. (2002). The chemical organization of signaling interactions. *Bioinformatics* 18, 855–863.

Bhalla, U. S. (2003). Understanding complex signaling networks through models and metaphors. *Prog. Biophys. Mol. Biol.* 81, 45–65.

Bhalla, U. S. (2014a). Molecular computation in neurons: a modeling perspective. *Curr. Opin. Neurobiol.* 25, 31–37. doi:10.1016/j.conb.2013.11.006.

Bhalla, U. S. (2014b). Multiscale modeling and synaptic plasticity. *Prog. Mol. Biol. Transl. Sci.* 123, 351–386. doi:10.1016/B978-0-12-397897-4.00012-7.

Bhalla, U. S. (2017). Synaptic input sequence discrimination on behavioral timescales mediated by reaction-diffusion chemistry in dendrites. *eLife* 6. doi:10.7554/eLife.25827.

Bhalla, U. S., and Iyengar, R. (1999). Emergent properties of networks of biological signaling pathways. *Science* 283, 381–387.

Bhalla, U. S., and Iyengar, R. (2001). Functional modules in biological signalling networks. *Novartis Found. Symp.* 239, 4–13; discussion 13-15, 45–51.

Bienenstock, E. L., Cooper, L. N., and Munro, P. W. (1982). Theory for the development of neuron selectivity: orientation specificity and binocular interaction in visual cortex. *J. Neurosci.* 2, 32–48.

Clarke, Bruce L. (1981). Complete set of steady states for the general stoichiometric dynamical system. *J. Chem. Phys.* 75, 4970–4979.

Dan, Y., and Poo, M.-M. (2004). Spike timing-dependent plasticity of neural circuits. *Neuron* 44, 23–30.

Danø, S., Madsen, M. F., Schmidt, H., and Cedersund, G. (2006). Reduction of a biochemical model with preservation of its basic dynamic properties. *FEBS J.* 273, 4862–4877. doi:10.1111/j.1742-4658.2006.05485.x.

Dormanns, K., Brown, R. G., and David, T. (2016). The role of nitric oxide in neurovascular coupling. *J. Theor. Biol.* 394, 1–17. doi:10.1016/j.jtbi.2016.01.009.

Frey, U., and Morris, R. G. (1997). Synaptic tagging and long-term potentiation. *Nature* 385, 533–536. doi:10.1038/385533a0.
Hayer, A., and Bhalla, U. S. (2005). Molecular switches at the synapse emerge from receptor and kinase traffic. *PLoS Comput. Biol.* 1, 137–154. doi:10.1371/journal.pcbi.0010020.
Higgins, D., Graupner, M., and Brunel, N. (2014). Memory maintenance in synapses with calcium-based plasticity in the presence of background activity. *PLoS Comput. Biol.* 10, e1003834. doi:10.1371/journal.pcbi.1003834.
Hudmon, A., Lebel, E., Roy, H., Sik, A., Schulman, H., Waxham, M. N., et al. (2005). A mechanism for Ca2+/calmodulin-dependent protein kinase II clustering at synaptic and nonsynaptic sites based on self-association. *J. Neurosci. Off. J. Soc. Neurosci.* 25, 6971–83. doi:10.1523/JNEUROSCI.4698-04.2005.
Jain, P., and Bhalla, U. S. (2009). Signaling logic of activity-triggered dendritic protein synthesis: an mTOR gate but not a feedback switch. *PLoS Comput. Biol.* 5, e1000287. doi:10.1371/journal.pcbi.1000287.
Kholodenko, B. N. (2000). Negative feedback and ultrasensitivity can bring about oscillations in the mitogen-activated protein kinase cascades. *Eur. J. Biochem.* 267, 1583–1588. doi:10.1046/j.1432-1327.2000.01197.x.
Kim, B., Hawes, S. L., Gillani, F., Wallace, L. J., and Blackwell, K. T. (2013). Signaling pathways involved in striatal synaptic plasticity are sensitive to temporal pattern and exhibit spatial specificity. *PLoS Comput. Biol.* 9, e1002953. doi:10.1371/journal.pcbi.1002953.
Lisman, J. (1989). A mechanism for the Hebb and the anti-Hebb processes underlying learning and memory. *Proc. Natl. Acad. Sci. U. S. A.* 86, 9574–9578.
Lomnitz, J. G., and Savageau, M. A. (2016). Design Space Toolbox V2: Automated Software Enabling a Novel Phenotype-Centric Modeling Strategy for Natural and Synthetic Biological Systems. *Front. Genet.* 7, 118. doi:10.3389/fgene.2016.00118.
Markram, H., Muller, E., Ramaswamy, S., Reimann, M. W., Abdellah, M., Sanchez, C. A., et al. (2015). Reconstruction and Simulation of Neocortical Microcircuitry. *Cell* 163, 456–492. doi:10.1016/j.cell.2015.09.029.
Maurya, M. R., Bornheimer, S. J., Venkatasubramanian, V., and Subramaniam, S. (2005). Reduced-order modelling of biochemical networks: application to the GTPase-cycle signalling module. *IEE Proc. Syst. Biol.* 152, 229–42.
Mayer, J., Khairy, K., and Howard, J. (2010). Drawing an elephant with four complex parameters. *Am. J. Phys.* 78, 648–649. doi:10.1119/1.3254017.
Novák, B., and Tyson, J. J. (2004). A model for restriction point control of the mammalian cell cycle. *J. Theor. Biol.* 230, 563–79.
Oliveira, R. F., Terrin, A., Di Benedetto, G., Cannon, R. C., Koh, W., Kim, M., et al. (2010). The role of type 4 phosphodiesterases in generating microdomains of cAMP: large scale stochastic simulations. *PloS One* 5, e11725. doi:10.1371/journal.pone.0011725.
Quaiser, T., Dittrich, A., Schaper, F., and Mönnigmann, M. (2011). A simple work flow for biologically inspired model reduction - application to early JAK-STAT signaling. *BMC Syst. Biol.* 5, 30. doi:10.1186/1752-0509-5-30.
Radulescu, O., Gorban, A. N., Zinovyev, A., and Lilienbaum, A. (2008). Robust simplifications of multiscale biochemical networks. *BMC Syst. Biol.* 2, 86. doi:10.1186/1752-0509-2-86.
Roelfsema, P. R., and Holtmaat, A. (2018). Control of synaptic plasticity in deep cortical networks. *Nat. Rev. Neurosci.* 19, 166–180. doi:10.1038/nrn.2018.6.

Savageau, M. A. (2001). Design principles for elementary gene circuits: Elements, methods, and examples. *Chaos* 11, 142–159. doi:10.1063/1.1349892.
Shouval, H. Z., Bear, M. F., and Cooper, L. N. (2002). A unified model of NMDA receptor-dependent bidirectional synaptic plasticity. *Proc. Natl. Acad. Sci. U. S. A.* 99, 10831–10836. doi:10.1073/pnas.152343099.
Singh, D., and Bhalla, U. S. (2018). Subunit exchange enhances information retention by CaMKII in dendritic spines. *eLife* 7. doi:10.7554/eLife.41412.
Smolen, P., Baxter, D. A., and Byrne, J. H. (2012). Molecular Constraints on Synaptic Tagging and Maintenance of Long-Term Potentiation: A Predictive Model. *PLoS Comput Biol* 8, e1002620. doi:10.1371/journal.pcbi.1002620.
Snowden, T. J., van der Graaf, P. H., and Tindall, M. J. (2017). Methods of Model Reduction for Large-Scale Biological Systems: A Survey of Current Methods and Trends. *Bull. Math. Biol.* 79, 1449–1486. doi:10.1007/s11538-017-0277-2.
Sorokina, O., Sorokin, A., and Armstrong, J. D. (2011). Towards a quantitative model of the post-synaptic proteome. *Mol. Biosyst.* 7, 2813–2823. doi:10.1039/c1mb05152k.
Stiles, J. R., and Bartol, T. M. (2001). “Monte Carlo methods for simulating realistic synaptic microphysiology using MCell.,” in *Computational Neuroscience: Realistic Modeling for Experimentalists, editor Erik De Schutter* (CRC Press), 87–127.
Südhof, T. C. (2017). Synaptic Neurexin Complexes: A Molecular Code for the Logic of Neural Circuits. *Cell* 171, 745–769. doi:10.1016/j.cell.2017.10.024.
Taylor, S. R., Doyle, F. J., and Petzold, L. R. (2008). Oscillator model reduction preserving the phase response: application to the circadian clock. *Biophys. J.* 95, 1658–1673. doi:10.1529/biophysj.107.128678.
Viswan, N. A., HarshaRani, G. V., Stefan, M. I., and Bhalla, U. S. (2018). FindSim: A Framework for Integrating Neuronal Data and Signaling Models. *Front. Neuroinformatics* 12. doi:10.3389/fninf.2018.00038.
Wallace, E. W. J., Gillespie, D. T., Sanft, K. R., and Petzold, L. R. (2012). Linear noise approximation is valid over limited times for any chemical system that is sufficiently large. *IET Syst. Biol.* 6, 102–115. doi:10.1049/iet-syb.2011.0038.
Wils, S., and De Schutter, E. (2009). STEPS: Modeling and Simulating Complex Reaction-Diffusion Systems with Python. *Front. Neuroinformatics* 3, 15. doi:10.3389/neuro.11.015.2009.
Wilson, D. E., Whitney, D. E., Scholl, B., and Fitzpatrick, D. (2016). Orientation selectivity and the functional clustering of synaptic inputs in primary visual cortex. *Nat. Neurosci.* 19, 1003–1009. doi:10.1038/nn.4323.
Zhu, C., Byrd, R. H., Lu, P., and Nocedal, J. (1997). Algorithm 778: L-BFGS-B: Fortran subroutines for large-scale bound-constrained optimization. *ACM Trans. Math. Softw.* 23, 550–560. doi:10.1145/279232.279236.