

A Neural Hybrid-System Model of the Basal Ganglia

Joseph Gerard Makin
Alessandro Abate



Electrical Engineering and Computer Sciences
University of California at Berkeley

Technical Report No. UCB/Eecs-2007-16

<http://www.eecs.berkeley.edu/Pubs/TechRpts/2007/Eecs-2007-16.html>

January 16, 2007

Copyright © 2007, by the author(s).
All rights reserved.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission.

A Neural Hybrid-System Model of the Basal Ganglia

Joseph G. Makin, Alessandro Abate

Department of Electrical Engineering and Computer Sciences
University of California, at Berkeley – Berkeley, CA 94720
{makin,aabate}@eecs.berkeley.edu

Abstract—The basal ganglia (BG) are a set of functionally related and structurally interconnected nuclei in the human brain which form part of a closed loop between cortex and thalamus, receiving input from the former and outputting to the latter. The BG have been implicated in motor control and cognitive switching tasks; in particular, it is believed that the BG function as a controller for motor tasks by selectively disinhibiting appropriate portions of the thalamus and hence activating, via a feedback loop, cortical regions. These switching behaviors are perforce discrete, whereas the underlying dynamics of neuron voltages and neurotransmitter levels are continuous-time, continuous-state phenomena. To this end, we propose and simulate a hybrid automaton for modeling individual neurons that affords explicit representation of voltage discharges and discrete outputs along with continuous voltage dynamics within a single, elegant model; and which is amenable both to the construction of large networks—in particular the cortico-basalthalamic loops—and to analysis on such networks.

I. INTRODUCTION

Motor control in the human brain has long been known to involve, *inter alia*, a series of topologically interconnected nuclei known collectively as the basal ganglia (BG). Specifically, these nuclei are believed to perform task-switching—for motor as well, perhaps, as cognitive routines—by first compressing cortical input and then selectively disinhibiting the appropriate region(s) of the thalamus. The disinhibited thalamic regions in turn activate corresponding regions of the cortex via feedback loops [1], [2]. This form of control has been likened to “releasing the brakes” on certain regions of the thalamus and hence enabling cortical activation [3]. These cortico-basal-thalamic loops are thought to be organized into parallel, segregated channels [1], [4].

Recently, the BG have attracted the attention of control theorists and other mathematical modelers, although at varying levels of model detail. These models can be divided roughly into two types: “Top-down” models take the BG network architecture as the point of departure either to implement a particular control scheme or to demonstrate some other systems-level property of the network. The results are then compared for congruence with, for example, properties of specific neuron populations. Alternatively, “bottom-up” schemes are simulations which attempt to reproduce the (putative) functionality of the BG by assembling models of neuron dynamics into large networks in accordance with BG connectivity.

So, for example, [4] has proposed that the BG implement a static and dynamic state (SDS) feedback control scheme

which performs speed-field tracking. This top-down scheme requires for the stability of the system that certain matrices in the equations of motion for the plant (i.e., the arm, leg, or etc. which is being controlled) be uniformly positive definite over the state space (“sign proper”). It is hypothesized that the parallel channels of the BG which are activated correspond to those which are sign-proper for the particular task at hand.

Alternatively, [5] begins with the premise that the BG perform action selection via signal selection, and then devises a neural network architecture for this purpose which is consistent with the architecture of the BG. Again, [6] has proposed that the BG instantiate an actor-critic system that implements reinforcement learning. Or again, [2] has modeled the BG using (modified) stochastic Petri nets, with particular attention to the firing functions for various nuclei and the fine-grained connectivity of the network.

All of these top-down models aim primarily at fidelity to the overall network architecture of the BG while implementing this or that computational scheme. The alternative is to build models from the bottom up, based on individual neural dynamics. This approach has been taken with the medium spiny neurons of the neostriatum [7]; the neurons of the subthalamic nucleus and the globus pallidus [8] (see Section II on the nuclei of the BG); and with a small number of loops from the prefrontal cortex through the direct pathway of the BG (see Section II), through the thalamus, and back to the cortex [9], [3]. These approaches share a focus on simulation rather than analysis, and quite generally aim at reproducing network-level behavior.

In the present paper, we propose a model situated between the top-down and bottom-up approaches, one which offers facility in simulation, susceptibility to analysis (particularly reachability proofs), and fidelity to the underlying neural dynamics. The model shares this much in common with those of Gillies ([10], [11]), but whereas the latter use non-linear dynamics which limit the feasibility of analysis, the present model proposes that individual neurons (or neural populations) be treated as hybrid systems. This provides in particular for (1) the reset-like nature of neuron voltage discharges; (2) discrete, on-or-off output, which appears to be a key feature of the BG (see Section II); and (3) a linear model of the system.

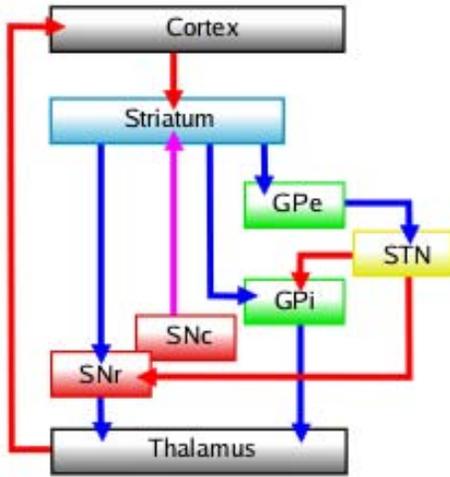


Fig. 1. The primary interconnections of the basal ganglia. Red arrows are glutaminergic (excitatory), blue are GABAergic (inhibitory), and magenta is dopaminergic (courtesy of Andrew Gillies).

II. OVERVIEW OF THE BASAL GANGLIA

The basal ganglia comprise four distinct nuclei: the striatum (STR), the globus pallidus, the subthalamic nucleus (STN), and the substantia nigra. The globus pallidus is further subdivided into an internal and an external segment, the GPi and GPe, respectively; and the substantia nigra divided into the pars reticula (“netlike”) or SNr, and the pars compacta or SNc [1].

The primary interconnections of the BG are shown in Fig. 1. Note that the SNc is a sort of controller, which mediates output of the STR via dopaminergic input. The STR can be thought of as the input nucleus of the BG, though it in turn receives input from the cortex. The GPi and SNr are the output nuclei of the BG: they tonically inhibit the thalamus, which in turn feeds back to the cortex. All the interconnections of the BG are in fact inhibitory except for the dopaminergic output of the SNc, which can be either excitatory or inhibitory; and the excitatory glutaminergic output of the STN [1], [11].

For our purposes, the following facts about the BG should also be noted:

- 1) The paths shown in Fig. 1 are not single channels, but series of segregated parallel channels (i.e., “topographically” organized). The entire network is vastly parallel [2].
- 2) The striatum has on the order of 10^8 cells, whereas the rest of the BG nuclei have on the order of 10^5 cells. This suggests that the BG perform some sort of data compression from the cortex [2].
- 3) The BG are widely believed to perform a sort of action selection by activating only certain of the topographically connected loops through the thalamus and back to the cortex [2], [1], [5]. Since in particular some of these actions are mutually exclusive with one another, this selection is an all-or-nothing affair; thus it is neither plausible nor, in many cases, possible to have actions

which have not been selected nevertheless be activated by however small a signal. So, e.g., since motor commands are produced directly in the motor and premotor regions of the cortex, these areas should receive only “go” and “no go” input from the thalamus; i.e. a discrete input.

- 4) The BG consist of two pathways: the direct pathway from STR through the output nuclei (SNr and GPi) to the thalamus; and the indirect pathway from the STR through the GPe to the STN and thence to the output nuclei. These two pathways have opposite effects on the thalamus: activation of the direct pathway serves to disinhibit the thalamus, whereas activation of the indirect pathway reduces inhibition of the thalamus. However, these two pathways in general work in concert, since the same mechanism which serves to activate the direct pathway deactivates the indirect pathway. This is the input from the SNc, the “controller.” Dopamine from the SNc facilitates firing of the so-called D1 receptors in the STN, which innervate the direct pathway, just as it reduces firing in the D2 receptors which feed the indirect pathway. Thus, we see that the two pathways work together by a “push-pull” mechanism: dopaminergic input to STR “pushes” the output nuclei via the excitatory influence of the indirect pathway, while the direct pathway remains quiescent; and reduction of SNc input “pulls back” the output nuclei via the inhibitory input of the direct pathway, while the indirect pathway stays inactive. This mechanism implements the action selection lately noted.
- 5) Several well-known disorders are caused by malfunction of the BG, viz. hyperkinetic disorders like Huntington disease and hemiballismus, and hypokinetic disorders, which includes Parkinson disease. Much is known about the pathologies of these diseases, and the clinical data enforce constraints on any computational model. So, for example, Parkinson disease is associated with a decrease in dopamine product, as well as overactivity in the indirect pathway. In particular, the hypoactivity is believed to result from increased output from the STN, which in turn increases inhibitory output from the GPi to the thalamus, decreasing activation of the cortex [1]. Similar observations have been made about other disorders of the BG.
- 6) Neurons in the striatum exhibit bistable behavior, meaning that sufficient upstream input can put the neuron into an “up” state, in which the neuron requires less input in order to fire.

III. THE MODEL

A. Constraints

The primary constraints on the model are as follows:

- *Scalability*. Since the current model assumes a bottom-up approach, the primitives must be simple enough that a network consisting of scores of them be simulatable. The

BG themselves consist of about 100 million neurons, so ideally the simulation could handle a network consisting of on the order of a million or more of the neuron models (it is supposed that every pathway in the BG need not be simultaneously simulated).

- *Amenability to analysis.* Unlike other bottom-up models, the present approach is intended to be susceptible to analysis in addition to simulation. We are particularly interested in reachability analyses; so, for example, that it might be shown that overactivity (specified in terms of mean firing rate) of the GPi results from (is reachable from) a certain range of dopamine upstream in the STR.
- *Discrete Output.* As we saw in Section II, the basal ganglia implement discrete switching behaviors, which means that output from neurons should be cast as a simple “on” or “off.” There are other approaches to capturing this switching behavior, among them a winner-take-all approach [5], but there is little biological evidence for such a mechanism in the BG; whereas the present proposal is consistent with the evidence that neuronal output which does not exceed a certain threshold is simply ignored by the downstream neuron. Additionally, modeling neurons as either on or off lends itself nicely to modeling the up and down states of the striatal medium spiny neurons. Finally, on/off output is much simpler for both analysis and simulation than winner-take-all implementations.
- *Fidelity to neural dynamics.* [2] has constructed a vastly parallel network that is consistent with constraints (1)-(3), but relies on a purely discrete formalism which ignores the underlying neural dynamics. We hypothesize that such dynamics are crucial to capturing the functionality of the BG.

B. Structure of the Model

The hybrid automaton is depicted in Fig. 2. The first equation governs the membrane voltage, which functions like a capacitor (hence the proportionality to V). An input vector \mathbf{y} consists of the output of other neurons. Note that this vector \mathbf{y} is not the same as the scalar y which appears in the fourth differential equation; the latter is the output of the present neuron, whereas the former contains the outputs of all the neurons which input to the present one. These are weighted according to some parameter vector \mathbf{a} (hence the inner product).

Next consider the equation for $z(t)$, which is simply the spike train $x(t)$ convolved with a Heaviside function minus another delayed (by W seconds) Heaviside function. Thus $z(t)$ is a positive integer-valued function which simply counts the number of spikes that occur within a moving period W . This captures the fact that neural signals use frequency-coding over short ($W = 100$ ms) windows.

The spike train is produced by a reset map triggered by the voltage. Note that all reset maps are taken to be forcing rather than enabling; the automaton was depicted thus rather than using domains to enforce transitions for the sake of simplicity. When the membrane voltage exceeds a certain threshold θ , it

is reset to its resting potential and a spike emitted from the axon. Thus $x(t)$ is reset to a Dirac delta at the time t of threshold crossing, and the voltage is reset to zero. Note that the resting potential of a neuron is in fact about -70 mV, but the model simply adds a DC shift for simplicity; the threshold is similarly shifted, so the net effect on the model is nil.

The two remaining reset maps turn the neuron “on” and “off”; that is, they encode whether the neural spiking frequency is sufficient to count as a signal to downstream neurons. If this threshold, α , is exceeded while the neuron is in its off state, then the output y is turned on; the same applies *mutatis mutandis* to switching from on to off. This same mechanism may be used (though it is not depicted in Fig. 2) to model bistable neurons (see Section II): the on (off) reset map, in addition to toggling the output y , also resets the threshold θ to a lower (higher) value. This captures the increased susceptibility to input evinced by such neurons in their “up” state.

Finally, both the output y and the spiking x have trivial (i.e. no) dynamics associated with them, hence the final two equations of the automaton are zero-valued.

C. Dynamics of the Hybrid Automaton

The hybrid automaton proposed in the previous section can be rewritten in the following form:

$$\begin{pmatrix} \dot{v}(t) \\ \dot{z}(t) \\ \dot{x}(t) \end{pmatrix} = A_1 \begin{pmatrix} v(t) \\ z(t) \\ x(t) \end{pmatrix} + A_2 \begin{pmatrix} v(t-W) \\ z(t-W) \\ x(t-W) \end{pmatrix} + B \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_M \end{pmatrix} :$$

$$A_1 = \begin{pmatrix} -\tau & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}, \quad A_2 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 0 & 0 \end{pmatrix},$$

$$B = \begin{pmatrix} a_1 & a_2 & \dots & a_M \\ 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \end{pmatrix} \quad (1)$$

$$y = \begin{cases} 1 & , z(t) \geq \alpha \\ 0 & , z(t) < \alpha \end{cases} \quad (2)$$

where M is the number of inputs to the system. The inputs, $\{y_i\}_{i=1}^M$, are the outputs of the upstream neurons. The reset maps are not listed here, but are just as given in Section III-B. Note in particular that there are no continuous dynamics associated with $x(t)$; the dynamics are rather given entirely by the resets.

Equation (1) is a linear delay-differential equation (DDE), and as such is amenable to reachability analysis via the techniques for linear DDEs (though we will avoid these complications – see below). This model is also perfectly appropriate to the bistable neurons of the striatum: as lately noted, the reset map on the output need only be modified to additionally lower or raise the membrane threshold voltage (θ) according to whether the output is on or off, respectively.

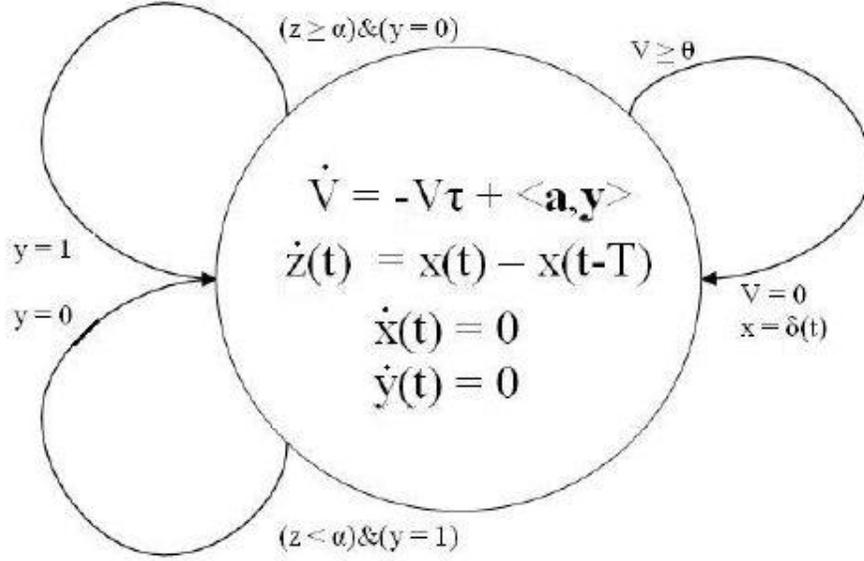


Fig. 2. A hybrid automaton representing neural dynamics. There are three state variables, an output y , and three reset maps.

D. An Approximate Model

We began by proposing a state variable z that was a convolution of a spike train with a “time window,” i.e. the difference between a Heaviside function and another delayed (by W seconds) Heaviside function:

$$z(t) = x(t) * h(t) \quad (3)$$

where

$$h(t) := u(t) - u(t - W) \quad (4)$$

is the difference of Heaviside functions (unit steps). Recall that equations (3) and (4) can be written as the differential equation which appears in the state-space equation (1):

$$\dot{z} = x(t) - x(t - W). \quad (5)$$

Now, the reset maps entail that $x(t)$ is a spike train, i.e. a sum over Dirac deltas with various delays τ_i :

$$x(t) := \sum_{i=1}^I \delta(t - \tau_i). \quad (6)$$

Writing out the convolution explicitly in terms of an integral, we can see that $z(t)$ is nothing but a count of the number of spikes that have occurred in the preceding W seconds:

$$\begin{aligned} z(t) &= \int_{-\infty}^{+\infty} x(s)[u(t-s) - u(t-W-s)]ds \\ &= \int_{-\infty}^{+\infty} x(s)u(t-s)ds - \int_{-\infty}^{+\infty} x(s)u(t-W-s)ds \\ &= \int_{-\infty}^t x(s)ds - \int_{-\infty}^{t-W} x(s)ds \\ &= \sum_{i=1}^I \int_{t-W}^t \delta(s - \tau_i)ds. \end{aligned} \quad (7)$$

Unfortunately, eq. (5) is a delay differential equation (DDE), which renders reachability analysis rather complicated. Our model is also so far strictly deterministic, whereas neural firing is more accurately conceived of as a stochastic process. However, a single felicitous substitution can simultaneously resolve both of these difficulties.

Consider first substitution of eq. (6) into the first line of (7):

$$z(t) = \sum_{i=1}^I \int_{-\infty}^{+\infty} \delta(s - \tau_i)[u(t-s) - u(t-W-s)]ds. \quad (8)$$

Now employing the change of variables $\lambda_i = s - \tau_i$ for each i yields

$$\begin{aligned} z(t) &= \sum_{i=1}^I \int_{-\infty}^{+\infty} \delta(\lambda_i)[u(t - \lambda_i - \tau_i) - u(t - W - \lambda_i - \tau_i)]d\lambda_i \\ &= \sum_{i=1}^I \left[\int_{-\infty}^{t-\tau_i} \delta(\lambda_i)d\lambda_i - \int_{-\infty}^{t-\tau_i-W} \delta(\lambda_i)d\lambda_i \right] \\ &= \sum_{i=1}^I \left[\int_{-\infty}^{t-\tau_i} \delta(\lambda_i)d\lambda_i - \int_{-\infty}^{t-\tau_i} \delta(\eta_i - W)d\eta_i \right] \end{aligned} \quad (9)$$

where the last line follows by changing variables again, this time in the second term and via $\eta_i = \lambda_i + W$.

We now “relax” our spike counter by substituting a probability density function for each impulse function in eq. (9). Our choice of pdf is the Erlang density function

$$f_{Erlang}(t|n, \alpha) = \frac{\alpha^n t^{n-1}}{(n-1)!} e^{-\alpha t}, \quad (10)$$

since in the limit as n and α approach $+\infty$, this function approaches a Dirac delta situated at the mean, $\frac{n}{\alpha}$:

$$\lim_{n, \alpha \rightarrow \infty} \frac{\alpha^n t^{n-1}}{(n-1)!} e^{-\alpha t} = \delta\left(t - \frac{n}{\alpha}\right). \quad (11)$$

(This can be seen by noting that the variance of the Erlang distribution is $\frac{n}{\alpha^2}$. If the mean is kept constant while n and α are increased toward ∞ , then the variance will approach zero.) Reconsidering eq. (9) in light of this pdf, we find that

$$\begin{aligned} z(t) &= \sum_{i=1}^I \left[\int_{-\infty}^{t-\tau_i} f_E(\lambda_i | n_1, \alpha_1) d\lambda_i - \int_{-\infty}^{t-\tau_i} f_E(\eta_i | n_2, \alpha_2) d\eta_i \right], \\ &= \sum_{i=1}^I \left[P_E(T \leq t - \tau_i | n_1, \alpha_1) - \right. \\ &\quad \left. P_E(T \leq t - \tau_i | n_2, \alpha_2) \right], \end{aligned} \quad (12)$$

where, finally, we choose

$$\begin{aligned} \frac{n_1}{\alpha_1} &= 0 \\ \frac{n_2}{\alpha_2} &= W \end{aligned} \quad (13)$$

How shall we interpret this substitution? Let us first consider the meaning of the Erlang distribution. Suppose that the process of interest is Poisson. (That is, if we divide the time interval over which the process occurs into subintervals, then the following three properties hold: [1] the probability of an event occurring in any one subinterval is the same as the probability of an event occurring in any other subinterval; [2] the probability of more than one event occurring in a subinterval is zero; and [3] events in one subinterval are independent of events in other subintervals.) For such a process, we may want to know the probability that exactly n events (an integer number) have occurred in a certain period of time, given the rate at which events take place, α . This probability is given by the Erlang probability density function (10).

Thus the value of z at time t is a sum of the probabilities associated with I different Poisson processes, each one representing a neural discharge (the first cumulative density function) and a “forgetting” of that discharge (the second cdf), i.e. its passing out of the time window of the neuron. Thus i^{th} term is the difference of two probabilities: The first is the probability that n_1 events have occurred in the i^{th} process by time $t - \tau_i$, given that the process started at time $t = 0$ and that there are α_1 events/second. Equivalently, this density function represents the probability that n_1 events have occurred in the i^{th} process by time t , given that the process started at time τ_i , and again that events occur at the rate α_1 events/second. Since τ_i is the time of the i^{th} threshold crossing, this may be interpreted as the probability that the n_1 independent events necessary for neural discharge (“firing”) have occurred in the T seconds following the voltage threshold crossing. The mean time for T is $\frac{n_1}{\alpha_1} = 0$, from eq. (13).

The second of the two probabilities in the difference given by the i^{th} term is just like the first, save that the number of events and rate are n_2 and α_2 , respectively, and that consequently the mean time for the occurrence of the n_2^{th} event is W seconds (from eq. (13)) after the starting time, τ_i .

So: we can interpret substitution of Erlang density functions for Dirac deltas as a transformation of the original deterministic system into a stochastic one. Neural firings are no

longer deterministically occasioned by threshold crossings but now have their expectation value at the threshold crossing—hence the mean at τ_i . The subtracted probability on the other hand represents the “forgetting” of past threshold crossings (i.e. their passing out of the moving time window), which occurs at a mean time of W seconds after the original crossing. The firings minus the “forgotten” firings are summed over all the spikes that have occurred; this is exactly what eq. (12) represents. And of course, as we let n and α approach $+\infty$ while maintaining the value of the ratios, the variances of both probabilities approach 0 and eq. (12) reduces to eq. (7), a deterministic running sum of the total number of spikes in the last W seconds.

We now consider what the substitution of the Erlang density function for the Dirac delta has bought us in the way of eliminating the DDE in eq. (5). Returning again to eq. (3), this time substituting in eqs. (6) and (4), yields

$$z(t) = \sum_{i=1}^I \delta(t - \tau_i) * [u(t) - u(t - W)]. \quad (14)$$

Taking Laplace transforms,

$$\begin{aligned} \hat{z}(s) &= \left(\sum_{i=1}^I e^{s\tau_i} \right) \left(\frac{1 - e^{sW}}{s} \right) \\ &= \left(\sum_{i=1}^I \frac{1}{s} e^{s\tau_i} \right) \left(1 - e^{sW} \right) \end{aligned} \quad (15)$$

so that transforming back,

$$\begin{aligned} z(t) &= \left(\sum_{i=1}^I u(t - \tau_i) \right) * [\delta(t) - \delta(t - W)] \\ &= \zeta(t) * [\delta(t) - \delta(t - W)], \end{aligned} \quad (16)$$

where

$$\zeta(t) := \sum_{i=1}^I u(t - \tau_i). \quad (17)$$

Replacing the Dirac deltas once more with the appropriate versions of the Erlang pdf from eq. (10) gives

$$z(t) = \zeta(t) * \left[\frac{\alpha_1^{n_1} t^{n_1-1}}{(n_1 - 1)!} e^{-\alpha_1 t} - \frac{\alpha_2^{n_2} t^{n_2-1}}{(n_2 - 1)!} e^{-\alpha_2 t} \right]. \quad (18)$$

Now, since the mean of the first distribution is zero (eq. (13)), we would like n_1 to be as small as possible, i.e. $n_1 = 1$ (we can’t be waiting for less than one event). We leave α_1 as a free parameter for now, but it is clear that we need it to be as large as possible to satisfy eq. (13), subject to our other constraints, which we shall consider shortly. The choice of n_2 as well is constrained by eq. (13) for the mean. However, as we shall see presently, n_2 will also determine the number of additional dimensions of the state space, so we would like to keep this parameter small. Finally, the Erlang density more closely approximates the impulse function as the variance $\frac{n_2}{\alpha_2^2}$ decreases. However, we cannot simply choose n_2 arbitrarily small (say, $n_2 = 1$) and α_2 arbitrarily large in order

to achieve this while keeping the state space small, since eq. (13) constrains the ratio of the two.

Bearing these considerations in mind and letting $n_1 = 1$, we look for a way to simplify the equation

$$z(t) = \zeta(s) * \alpha_1 e^{\alpha_1 t} - \zeta(s) * \frac{-\alpha_2^{n_2} t^{n_2-1}}{(n_2-1)!} e^{-\alpha_2 t}. \quad (19)$$

There is in fact a technique for reducing this equation to a simple set of linear ODEs, known as the ‘‘linear chain trick.’’ For simplicity we now let $\alpha_1 = \alpha_2 = \alpha$. The first step is to define a new state variable as

$$\xi_1(t) := \zeta(t) * \alpha e^{-\alpha t}, \quad (20)$$

which yields the differential equation

$$\dot{\xi}_1(t) = \alpha[\zeta(t) - \xi_1(t)], \quad (21)$$

or, in the Laplace domain,

$$\hat{\xi}_1(s) = \frac{\alpha}{s + \alpha} \hat{\zeta}(s), \quad (22)$$

where indeed $\hat{\xi}_1(s)$ denotes the Laplace transform of $\xi_1(t)$ and likewise $\hat{\zeta}(s)$ that of $\zeta(t)$. Similarly, we define a set of state variables $\{\hat{\zeta}_j(t)\}_{j=2}^{n_2}$ by

$$\xi_j(t) := \zeta(t) * \frac{\alpha^j t^{j-1}}{(j-1)!} e^{-\alpha t}. \quad (23)$$

We take Laplace transforms, factor out a term, and then substitute in the first line but at $j-1$:

$$\begin{aligned} \hat{\xi}_j(s) &= \frac{\alpha^j}{(s + \alpha)^j} \hat{\zeta}(s) \\ &= \frac{\alpha}{s + \alpha} \frac{\alpha^{j-1}}{(s + \alpha)^{j-1}} \hat{\zeta}(s) \\ &= \frac{\alpha}{s + \alpha} \hat{\xi}_{j-1} \end{aligned} \quad (24)$$

Performing the inverse transform yields the governing ODEs:

$$\dot{\xi}_j(t) = \alpha[\xi_{j-1}(t) - \xi_j(t)], \quad j = 2, 3, \dots, n_2. \quad (25)$$

Furthermore, it will be noticed from equations (19), (20), and (23) that

$$z(t) = \xi_1 - \xi_{n_2} \quad (26)$$

and hence from (21) and (25):

$$\dot{z}(t) = \alpha[\zeta(t) + \xi_{n_2}(t) - \xi_1(t) - \xi_{n_2-1}(t)]. \quad (27)$$

Finally, we let $\zeta(t)$ be a new state variable, with no dynamics; it is rather updated solely by the reset map:

$$\dot{\zeta}(t) = 0 \quad (28)$$

$$\begin{aligned} G(q) &= \{x \in \mathbb{R}^{n_2+3} | v \leq \theta\}, \\ R(q, \zeta) &= \zeta + 1, \end{aligned} \quad (29)$$

where G and R are the relevant guard and reset maps, respectively; v is the voltage; θ is the voltage threshold; q is the sole discrete state, and $x = (v, \xi_1, \dots, \xi_{n_2}, \zeta, z)^T$ is the

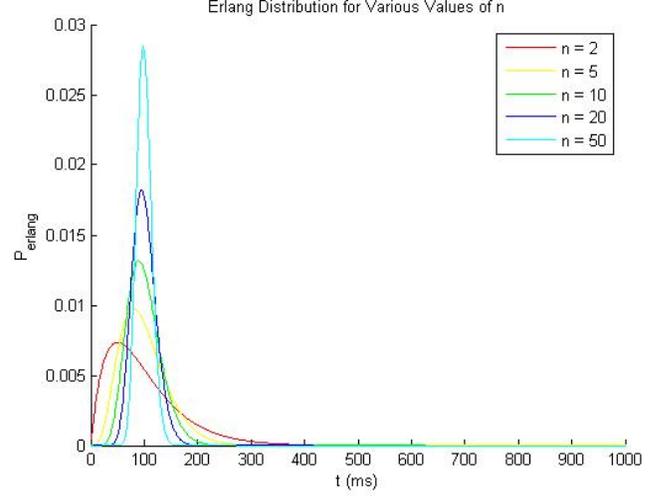


Fig. 3. Erlang distributions for various values of n and α . The ratio of the two was kept constant at $W = 100$ ms.

continuous state vector. We shall see presently why it is of length $n_2 + 3$.

We are at long last in position to rewrite the state-space equations, this time as a simple linear system:

$$\begin{pmatrix} \dot{v}(t) \\ \dot{\xi}_1(t) \\ \vdots \\ \dot{\xi}_{n_2}(t) \\ \dot{\zeta}(t) \\ \dot{z}(t) \end{pmatrix} = \begin{pmatrix} -\tau & 0 & 0 & \cdots & 0 & 0 & 0 & 0 \\ 0 & & & D & & \alpha & 0 & \\ 0 & & & & & 0 & 0 & \\ \vdots & & & & & \vdots & \vdots & \\ 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 \\ 0 & -\alpha & 0 & \cdots & -\alpha & \alpha & \alpha & 0 \end{pmatrix} \begin{pmatrix} v(t) \\ \xi_1(t) \\ \vdots \\ \xi_{n_2}(t) \\ \zeta(t) \\ z(t) \end{pmatrix} + \begin{pmatrix} a_1 & a_2 & \cdots & a_M \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_M \end{pmatrix}, \quad (30)$$

where

$$D = \begin{pmatrix} -\alpha & 0 & 0 & \cdots & 0 \\ \alpha & -\alpha & 0 & \cdots & 0 \\ 0 & \alpha & -\alpha & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \alpha & -\alpha \end{pmatrix}. \quad (31)$$

We are finally left with a choice for the value of n_2 which, as Fig. 3 shows, determines the shape of the approximant Erlang distribution.

IV. RESULTS

The hybrid automaton system of eq. (1) was simulated in MATLAB. A network of neurons was assembled to demonstrate a proof of concept and is not strictly faithful to the interconnections of the BG. Fig. 4 depicts the interconnections of this network. Note that the inputs were chosen (somewhat arbitrarily) to be sine waves. The architecture was chosen to correspond to a small-scale neural network in the BG: activation propagates uni-directionally and is organized roughly into layers.

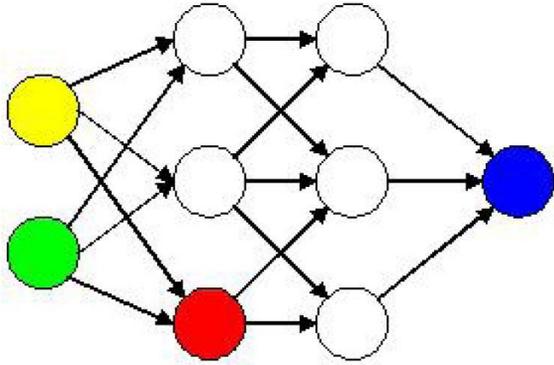


Fig. 4. The interconnections of the simulation. The colors correspond to the colors in Fig. 5. The green and yellow nodes are inputs (sine waves) and the blue is the output.

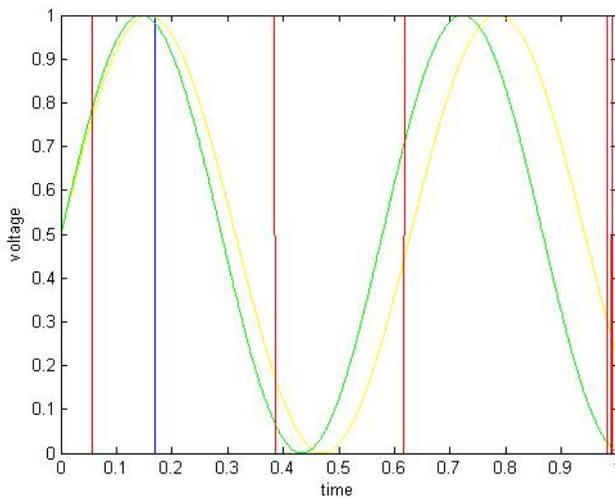


Fig. 5. The output of the input nodes (green and yellow), output node (blue), and a middle-layer node (red). The units of time and voltage are arbitrary.

Fig. 5 shows the neural output with time for the input nodes, output nodes, and a middle-layer node. Note that the network outputs a discrete signal, which can be interpreted as “go”/“no go” signal; this was one of our constraints for a BG simulation. Secondly, the execution time for this program, running on a 1.30 GHz machine with 256 MB of RAM, was 0.3 seconds. A similar model which is twice as large runs in about 0.5 seconds. If we assume that computation time is linear in number of units, then a network consisting of one million neurons—a reasonable approximation of the BG—will take about 12 hours, not an unreasonable figure.

V. FUTURE WORK

The hybrid automaton developed in this paper should now be assembled into a larger network which captures the connectivity of the basal ganglia. The functionality of the BG could then be simulated and specific hypotheses tested against the simulation. In particular, the disease pathologies mentioned in Section II must be reproducible in the model. The model could

then be used to generate further hypotheses for clinicians, whence the model parameters and interconnections be further refined. The second future direction, contingent on the first, is the analysis of the basal-ganglia network. Again, disease pathologies would provide the constraints; in this case, “downstream” manifestations of the disease (e.g. in the thalamus or cortex) would provide bounds for a backward reachability analysis to input or other parameters “upstream.” It should be stressed that the current model appears to be unique in providing the possibility of analysis of the BG at this level of detail.

VI. CONCLUSION

A hybrid-system model of neural dynamics was proposed for specific application in the basal ganglia, which exhibit discrete switching behavior, superimposed on continuous lowlevel dynamics. The proposed formalism lends itself both to simulation and analysis, though neither was performed in the present paper on large scale networks. Instead, the dynamics were simulated on a small scale, and the analysis procedure outlined. The present model meets the constraints of scalability, fidelity to neural dynamics, amenability to analysis, and discrete output (switching) behavior.

ACKNOWLEDGMENTS

The authors would like to thank Dr. Srin Narayanan and Aaron Ames for their helpful input.

REFERENCES

- [1] Eric R. Kandell, James H. Schwartz, and Thomas M. Jessell. *Principles of Neural Science*. McGraw-Hill Medical, 4th edition, 2000.
- [2] Srin Narayanan. The role of cortico-basal-thalamic loops in cognition: A computation model and preliminary results. *Neurocomputing*, 52-54:605–614, 2003.
- [3] Michael J. Frank, Bryan Loughry, and Randall C. O’Reilly. Interactions between the frontal cortex and the basal ganglia in working memory: A computational model. *Cognitive, Affective, and Behavioral Neuroscience*, 1(2):137–160, 2001.
- [4] A. Lőrincz. Static and dynamic state feedback control model of basal ganglia-thalamocortical loops. *International Journal of Neural Systems*, 8:339–357, 1997.
- [5] K. Gurney, T.J. Prescott, and P. Redgrave. A computational model for action selection in the basal ganglia. *Biological Cybernetics*, 84:401–410, 2001.
- [6] Andrew G. Barto. *Models of information processing in the basal ganglia*, chapter Adaptive critics and the basal ganglia, pages 215–232. MIT Press, Cambridge, MA, 1995.
- [7] John Randall Gobel. *University of California, San Diego*. PhD thesis, The role of the neostriatum in the execution of action sequences, 1997.
- [8] D. Terman, J.E. Rubin, A.C. Yew, and C.J. Wilson. Activity patterns in a model for the subthalamic network of the basal ganglia. *The Journal of Neuroscience*, 22(7):2963–2976, 2002.
- [9] Randall C. O’Reilly and Michael J. Frank. Making working memory work: a computational model of learning in the prefrontal cortex and basal ganglia. Technical report, Dep’t of Psychology, University of Colorado, Boulder, March 2005.
- [10] A.J. Gillies and D.J. Willshaw. A massively connected subthalamic nucleus leads to the generation of widespread pulses. *Proceedings. Biological Sciences/The Royal Society*, 265(1410):2101–2109, 1998.
- [11] A.J. Gillies, David Willshaw, and Zhaoping Li. Subthalamic-pallidal interactions are critical in determining normal and abnormal functioning of the basal ganglia. *Proceedings. Biological Sciences/The Royal Society*, 269(1491):545–551, 2002.