

# Low-dimensional, morphologically accurate models of subthreshold membrane potential

Anthony R. Kellems · Derrick Roos ·  
Nan Xiao · Steven J. Cox

Received: 25 June 2008 / Revised: 24 October 2008 / Accepted: 19 December 2008 / Published online: 27 January 2009  
© Springer Science + Business Media, LLC 2009

**Abstract** The accurate simulation of a neuron's ability to integrate distributed synaptic input typically requires the simultaneous solution of tens of thousands of ordinary differential equations. For, in order to understand how a cell distinguishes between input patterns we apparently need a model that is biophysically accurate down to the space scale of a single spine, i.e., 1  $\mu\text{m}$ . We argue here that one can retain this highly detailed input structure while dramatically reducing the overall system dimension if one is content to accurately reproduce the associated membrane potential at a small number of places, e.g., at the site of action potential initiation, under subthreshold stimulation. The latter hypothesis permits us to approximate the active cell model with an associated quasi-active model, which in turn we reduce by both time-domain (Balanced Truncation) and frequency-domain ( $\mathcal{H}_2$  approximation of the transfer function) methods. We apply and contrast these methods on a suite of typical cells, achieving up to four orders of magnitude in dimension reduction and an associated speed-up in the simulation of dendritic democratization and resonance. We also append a threshold mechanism and indicate that this reduction has the potential to deliver an accurate quasi-integrate and fire model.

**Keywords** Balanced truncation · Krylov methods · Quasi-active · Dendritic democratization · Dendritic resonance · Integrate-and-fire · Compartmental models

## 1 Introduction

Accurate delineation of the single cell map from synaptic input to membrane potential at the site of action potential initiation requires large, complex models. The size is dictated by synapse density and total dendritic length. With approximately one synapse per micron on a cell with 5 mm of dendrite we arrive at 5000 compartments. The complexity arises from both geometric tapering and nonuniform distribution of a variety of channel types. With both transient and persistent  $\text{Na}^+$  and  $\text{K}^+$  currents, a low threshold  $\text{Ca}^{2+}$  current, a leak current and current activated upon hyperpolarization we reach 10 (1 voltage and 9 gating) equations per compartment. Altogether, we arrive at a coupled system of 50,000 ordinary differential equations.

Wilfrid Rall was the first to build such models and the first to contemplate their systematic reduction. In particular, he showed how cells with symmetric geometry, receiving symmetric synaptic input, satisfying the “3/2” law at each branch point, and possessing passive membranes, could be mapped onto a single straight passive cable (Rall 1959). Although his assumptions were slightly relaxed by (Schierwagen 1989) and (Poznanski 1991), they remain too restrictive to be adopted by the general investigator of synaptic integration. If one is content to reproduce only spike shapes or rates then one can arrive at useful reduced

---

**Action Editor:** Wulfram Gerstner

---

A. R. Kellems (✉) · D. Roos · N. Xiao · S. J. Cox  
Department of Computational and Applied Mathematics,  
Rice University, Houston, TX, USA  
e-mail: tkellems@rice.edu

models by simply lumping compartments together. (Bush and Sejnowski 1993) and (Pinsky and Rinzel 1994) for example, achieve their ends with merely nine and two compartments, respectively. Both models of course surrender spatial precision of individual synapses.

We here employ old and new tools from the field of linear dynamical systems to reduce models with tens of thousands of equations to models with tens of equations without sacrificing either precise synaptic placement or accuracy of the associated subthreshold membrane potential at the site of action potential initiation.

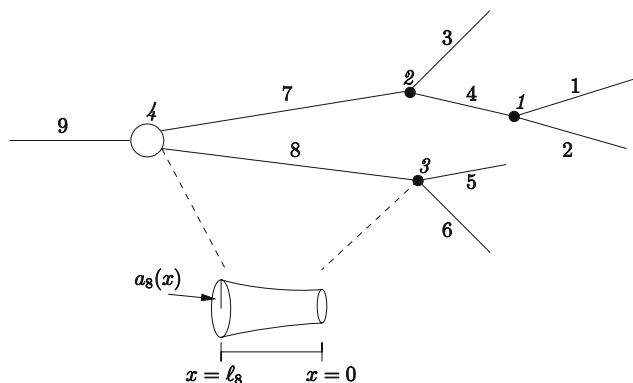
We begin, in Section 2, by specifying the full, nonlinear, nonuniform, branched, synaptically driven model. We compute the associated rest potential, linearize the full system about rest, and arrive at what Koch terms the quasi-active system (Koch 1999). In Section 3 we apply two distinct methods (one in the time domain, the other, in the frequency domain) of model reduction. They share the ability to reduce the dimension of the “internal” dynamics without compromising the input-output map. In Section 4 and Section 5 we test and contrast these two strategies via investigations of both dendritic democratization and dendritic resonance. In Section 6 we append a threshold mechanism to our reduced model and arrive at a morphologically accurate integrate-and-fire model. We close in Section 7 with a discussion of limitations and extensions.

## 2 The quasi-active system

We consider dendritic neurons (see, e.g., Fig. 1) with  $D$  branched dendrites meeting at the soma. The  $d$ th dendrite possesses  $B_d$  branches, and so the neuron possesses

$$B \equiv \sum_{d=1}^D B_d$$

branches. We denote by  $\ell_b$  the length of branch  $b$  and encode its radius, as a function of distance from its distal end, by  $a_b(x)$ . The transmembrane potential along branch  $b$  will be denoted by  $v_b(x, t)$ . We assume that the axial resistivity,  $R_i$  ( $\text{k}\Omega \text{ cm}$ ), and membrane capacitance,  $C_m$  ( $\mu\text{F}/\text{cm}^2$ ), are uniform throughout the cell. We suppose that branch  $b$  carries  $C$  distinct currents, with associated densities,  $G_{bc}(x)$  ( $\text{mS}/\text{cm}^2$ ) and reversal potentials  $E_c$ ,  $c = 1, \dots, C$ . The kinetics of current  $c$  on branch  $b$  are governed by (powers of) the  $F_c$  gating variables,  $w_{bcf}$ ,  $f = 1, \dots, F_c$ . When subjected to input



**Fig. 1** Example of a toy neuron to demonstrate our labeling scheme. Branches are indexed using the Hines ordering (Hines 1984), and junctions are labeled with italics. For this cell there are  $D = 3$  dendrites and  $J = 4$  junctions. Our labeling scheme implies, for example, that at junction 3 the children are indexed by  $b_3^1 = 5$ ,  $b_3^2 = 6$  and the mother is indexed by  $b_3^3 = 8$

at  $S_b$  synapses, these gating variables, together with  $v_b$ , obey the nonlinear cable equation

$$\begin{aligned} a_b C_m \partial_t v_b &= \frac{1}{2R_i} \partial_x (a_b^2 \partial_x v_b) \\ &\quad - a_b \sum_{c=1}^C G_{bc}(x) (v_b - E_c) \prod_{f=1}^{F_c} w_{bcf}^{q_{cf}} \\ &\quad - \frac{1}{2\pi} \sum_{s=1}^{S_b} g_{bs}(t) \delta(x - x_{bs}) (v_b - E_{bs}) \end{aligned} \quad (1)$$

$$\partial_t w_{bcf} = \frac{w_{cf,\infty}(v_b) - w_{bcf}}{\tau_{cf}(v_b)}, \quad 0 < x < \ell_b, \quad 0 < t. \quad (2)$$

Here  $g_{bs}(t)$  (nS) is the time course,  $x_{bs}$  is the spatial location, and  $E_{bs}$  is the reversal potential of the  $s$ th synapse on branch  $b$ . In this paper,  $E_{bs} = 0$  mV.

These branch potentials interact at  $J$  junction points, where junction  $J$  denotes the soma. At junction  $j < J$  we denote by  $b_j^1$  and  $b_j^2$  the branch indices of the two children and by  $b_j^3$  the branch index of the mother. Continuity of the potential at each junction requires

$$v_{b_j^1}(\ell_{b_j^1}, t) = v_{b_j^2}(\ell_{b_j^2}, t) = v_{b_j^3}(0, t), \quad j = 1, \dots, J-1 \quad (3)$$

while current balance there requires

$$\begin{aligned} a_{b_j^1}^2(\ell_{b_j^1}) \partial_x v_{b_j^1}(\ell_{b_j^1}, t) + a_{b_j^2}^2(\ell_{b_j^2}) \partial_x v_{b_j^2}(\ell_{b_j^2}, t) \\ = a_{b_j^3}^2(0) \partial_x v_{b_j^3}(0, t), \quad j = 1, \dots, J-1. \end{aligned} \quad (4)$$

Our  $D$  dendrites join at the soma and associated branch indices are  $b_j^d, d = 1, \dots, D$ . Continuity of the potential at the soma then reads

$$v_{b_j^1}(\ell_{b_j^1}, t) = v_{b_j^d}(\ell_{b_j^d}, t), \quad d = 2, \dots, D \quad (5)$$

while current balance there requires

$$\begin{aligned} C_m \partial_t v_{b_j^1}(\ell_{b_j^1}, t) = & \frac{\pi}{A_\sigma R_i} \sum_{d=1}^D a_{b_j^d}^2(\ell_{b_j^d}) \partial_x v_{b_j^d}(\ell_{b_j^d}, t) \\ & - \sum_{c=1}^C G_{\sigma c} (v_{b_j^1}(\ell_{b_j^1}, t) - E_c) \prod_{f=1}^{F_c} w_{\sigma cf}^{q_{cf}}(t) \\ & - \frac{1}{A_\sigma} \sum_{s=1}^{S_\sigma} g_{\sigma s}(t) (v_{b_j^1}(\ell_{b_j^1}, t) - E_{\sigma s}) \end{aligned} \quad (6)$$

$$\begin{aligned} \partial_t w_{\sigma cf}(t) = & \frac{w_{cf,\infty}(v_{b_j^1}(\ell_{b_j^1}, t)) - w_{\sigma cf}(t)}{\tau_{cf}(v_{b_j^1}(\ell_{b_j^1}, t))}, \\ & 0 < x < \ell_b, \quad 0 < t. \end{aligned} \quad (7)$$

where  $\sigma$  is the somatic index and  $A_\sigma$  is the surface area of the soma. Finally, we denote by  $\mathcal{L}$  the set of leaf indices, where a leaf is a branch with no children. We suppose that each leaf is sealed at its distal end, i.e.,

$$\partial_x v_b(0, t) = 0, \quad b \in \mathcal{L}. \quad (8)$$

The somatic voltage is thus  $v_\sigma(t) \equiv v_{b_j^1}(\ell_{b_j^1}, t)$ . Initially the neuron is at rest, implying that  $\partial_t v_b(x, 0) = 0$ . We permit each synapse to have a tonic conductance  $\bar{g}_{bs}$ , as in (Manwani and Koch 1999), which yields the initial condition  $g_{bs}(0) = \bar{g}_{bs}$ . We solve for the rest state and denote it by  $\bar{v}_b(x)$ , and similarly for the gating variables, which yields the initial conditions

$$v_b(x, 0) = \bar{v}_b(x) \quad (9)$$

$$w_{b cf}(x, 0) = \bar{w}_{b cf}(x) \equiv w_{cf,\infty}(\bar{v}_b(x)). \quad (10)$$

Consider Eq. (1) for branch  $b$  at its resting potential  $\bar{v}_b$ . Small perturbations to this rest state ought to be well-represented by linear approximations to the differential equations (Koch 1999). If  $g_{bs} = \bar{g}_{bs} + \varepsilon \tilde{g}_{bs}(x, t)$ , where  $\varepsilon$  is small, then the perturbed voltage and gating variables are assumed to be

$$v_b = \bar{v}_b + \varepsilon \tilde{v}_b + \mathcal{O}(\varepsilon^2)$$

$$w_{b cf} = \bar{w}_{b cf} + \varepsilon \tilde{w}_{b cf} + \mathcal{O}(\varepsilon^2).$$

Substituting these into Eq. (1), we construct a linearized model by solving for the perturbation terms  $\tilde{v}_b, \tilde{w}_{b cf}$  of order  $\varepsilon$ . After substitution we find

$$\begin{aligned} \varepsilon a_b C_m \partial_t \tilde{v}_b = & \varepsilon \frac{1}{2R_i} \partial_x (a_b^2 \partial_x \tilde{v}_b) \\ & - a_b \sum_{c=1}^C G_{bc}(x) (\bar{v}_b + \varepsilon \tilde{v}_b - E_c) \\ & \times \prod_{f=1}^{F_c} (\bar{w}_{b cf} + \varepsilon \tilde{w}_{b cf}(t))^{q_{cf}} \\ & - \frac{1}{2\pi} \sum_{s=1}^{S_b} (\bar{g}_{bs} + \varepsilon \tilde{g}_{bs}(t)) \delta(x - x_{bs}) \\ & \times (\bar{v}_b + \varepsilon \tilde{v}_b - E_{bs}) \end{aligned} \quad (11)$$

$$\varepsilon \partial_t \tilde{w}_{b cf}(t) = \frac{w_{cf,\infty}(\bar{v}_b + \varepsilon \tilde{v}_b) - (\bar{w}_{b cf} + \varepsilon \tilde{w}_{b cf})}{\tau_{cf}(\bar{v}_b + \varepsilon \tilde{v}_b)}. \quad (12)$$

The initial conditions are now

$$\tilde{v}_b(x, 0) = \tilde{w}_{b cf}(x, 0) = 0$$

while boundary conditions, because they are already linear, remain the same as in Eqs. (3), (4), (8). The soma conditions contain nonlinear terms, but they may be linearized in the same manner as shown here.

Expanding  $w_{cf,\infty}$  and  $\tau_{cf}$  in Eq. (12) in a Taylor series about  $\bar{v}_b$  yields the linearized equations for the gating variables

$$\partial_t \tilde{w}_{b cf}(t) = \frac{w'_{cf,\infty}(\bar{v}_b) \tilde{v}_b - \tilde{w}_{b cf}}{\tau_{cf}(\bar{v}_b)}.$$

In Eq. (11), if the product of gating variables is written as

$$F_{bc}(\varepsilon) = \prod_{f=1}^{F_c} (\bar{w}_{b cf} + \varepsilon \tilde{w}_{b cf}(t))^{q_{cf}}$$

then differentiating with respect to  $\varepsilon$  yields

$$\begin{aligned} F'_{bc}(\varepsilon) = & \sum_{p=1}^{F_c} \left[ q_{cp} (\bar{w}_{b cp} + \varepsilon \tilde{w}_{b cp})^{q_{cp}-1} \right. \\ & \times \left. \prod_{f=1, f \neq p}^{F_c} (\bar{w}_{b cf} + \varepsilon \tilde{w}_{b cf}(t))^{q_{cf}} \tilde{w}_{b cp} \right]. \end{aligned}$$

The order 0 and order  $\varepsilon$  terms of  $F_{bc}$  are

$$F_{bc}(0) = \prod_{f=1}^{F_c} \bar{w}_{b cf}^{q_{cf}} \quad \text{and}$$

$$F'_{bc}(0) = \sum_{p=1}^{F_c} \left( q_{cp} \bar{w}_{b cp}^{q_{cp}-1} \prod_{f=1, f \neq p}^{F_c} \bar{w}_{b cf}^{q_{cf}} \tilde{w}_{b cp} \right),$$

respectively. Hence the linearized ionic currents take the form

$$I_{\text{ionic}}(x, \tilde{v}_b, \tilde{w}_b) = \sum_{c=1}^C G_{bc}(x) (F_{bc}(0) \tilde{v}_b + F'_{bc}(0) (\bar{v}_b - E_c)).$$

The linearized synaptic input in Eq. (11) has the form

$$I_{\text{synaptic}}(t, x, \tilde{v}_b) = \frac{1}{2\pi} \sum_{s=1}^{S_b} (\tilde{g}_{bs}(t) (\bar{v}_b - E_{bs}) + \bar{g}_{bs} \tilde{v}_b) \times \delta(x - x_{bs}).$$

To complete the linearization process, we equate the terms of order  $\varepsilon$  and arrive at

$$\partial_t \tilde{v}_b = \mathcal{D}_b \tilde{v}_b + K_b(x) \tilde{v}_b + \sum_{c=1}^C \sum_{f=1}^{F_c} d_{bcf}(x) \tilde{w}_{bcf} + \eta_b$$

$$\partial_t \tilde{w}_{bcf}(t) = \phi_{bcf}(x) \tilde{v}_b + \psi_{bcf}(x) \tilde{w}_{bcf},$$

where

$$\mathcal{D}_b = \frac{1}{2C_m R_{ia_b}} \partial_x (a_b^2 \partial_x),$$

$$K_b(x) = -\frac{1}{C_m} \sum_{c=1}^C G_{bc}(x) F_{bc}(0) - \frac{1}{2\pi a_b C_m} \sum_{s=1}^{S_b} \bar{g}_{bs} \delta(x - x_{bs})$$

$$d_{bcf}(x) = -\frac{1}{C_m} G_{bc}(x) (\bar{v}_b - E_c) q_{cf} \bar{w}_{bcf}^{q_{cf}-1} \prod_{p=1, p \neq f}^{F_c} \bar{w}_{bcp}^{q_{cp}}$$

$$\eta_b = -\frac{1}{2\pi a_b C_m} \sum_{s=1}^{S_b} \tilde{g}_{bs}(t) \delta(x - x_{bs}) (\bar{v}_b - E_{bs})$$

$$\phi_{bcf}(x) = \frac{w'_{cf,\infty}(\bar{v}_b)}{\tau_{cf}(\bar{v}_b)}, \quad \psi_{bcf}(x) = -\frac{1}{\tau_{cf}(\bar{v}_b)}.$$

It is now apparent that this is a linear system for the  $b$ th branch, namely

$$\partial_t \begin{pmatrix} \tilde{v}_b \\ \tilde{w}_{b11} \\ \vdots \\ \tilde{w}_{b1F_1} \\ \tilde{w}_{b21} \\ \vdots \\ \tilde{w}_{b2F_2} \\ \vdots \\ \tilde{w}_{bC1} \\ \vdots \\ \tilde{w}_{bCF_C} \end{pmatrix} = \begin{pmatrix} \mathcal{D}_b + K_b & d_{b11} & d_{b12} & \dots & d_{bCF_C} \\ \phi_{b11} & \psi_{b11} & & & \\ \phi_{b12} & & \psi_{b12} & & \\ \vdots & & & \ddots & \\ \phi_{bCF_C} & & & & \psi_{bCF_C} \end{pmatrix} \begin{pmatrix} \tilde{v}_b \\ \tilde{w}_{b11} \\ \vdots \\ \tilde{w}_{b1F_1} \\ \tilde{w}_{b21} \\ \vdots \\ \tilde{w}_{b2F_2} \\ \vdots \\ \tilde{w}_{bC1} \\ \vdots \\ \tilde{w}_{bCF_C} \end{pmatrix} + \begin{pmatrix} \eta_b \\ 0 \\ \vdots \\ 0 \end{pmatrix},$$

or, more concisely,

$$\partial_t \mathbf{z}_b = \mathcal{Q}_b \mathbf{z}_b + \mathbf{u}_b. \quad (13)$$

We discretize the neuron in space by dividing each branch into  $\gamma_b = \text{ceil}(\ell_b/h)$  compartments, where  $h$  is some desired step size. The connectivity of the full morphology is encapsulated in the Hines matrix  $\mathbf{H}$ , which is the spatial discretization of each  $\mathcal{D}_b$  coupled with Eq. (3) to Eqs. (5) and (8) (Hines 1984). Using the Hines matrix imposes an outside-in ordering of branches and compartments, which leads to minimal fill-in for Gaussian Elimination (Hines 1984). If  $m$  and  $n$  denote, respectively, the number of gating variables per compartment and the total number of compartments, i.e.

$$m = \sum_{c=1}^C F_c \quad \text{and} \quad n = 1 + \sum_{b=1}^B \gamma_b$$

then the discretized system has  $N = n(m+1)$  variables. Restricting our attention to the  $b$ th branch, we compartmentalize the voltage by  $\tilde{v}_{:,b} \in \mathbb{R}^{\gamma_b}$ , and similarly the gating variables, to obtain

$$\partial_t \mathbf{z}_b = \mathbf{Q}_b \mathbf{z}_b + \mathbf{B}_b \mathbf{u}_b$$

where

$$\mathbf{z}_b = \begin{pmatrix} \tilde{v}_{:,b} \\ \tilde{w}_{:,b11} \\ \vdots \\ \tilde{w}_{:,bCF_C} \end{pmatrix} \in \mathbb{R}^{\gamma_b(m+1)}$$

$$\mathbf{Q}_b = \begin{pmatrix} \text{diag}(K_{:,b})d_{:,b11}d_{:,b12}\dots d_{:,bCF_C} \\ \phi_{:,b11} & \psi_{:,b11} \\ \phi_{:,b12} & \psi_{:,b12} \\ \vdots & \ddots \\ \phi_{:,bCF_C} & \psi_{:,bCF_C} \end{pmatrix} \in \mathbb{R}^{\gamma_b(m+1) \times \gamma_b(m+1)}$$

$$\mathbf{B}_b = \begin{pmatrix} I \\ 0 \end{pmatrix} \in \mathbb{R}^{\gamma_b(m+1) \times \gamma_b}, \quad \mathbf{I} \in \mathbb{R}^{\gamma_b \times \gamma_b}, \quad \mathbf{u}_b \in \mathbb{R}^{\gamma_b}.$$

For example, if branch  $b$  receives synaptic input at  $x_{b1}$  and  $x_{b2}$ , then we have

$$\mathbf{u}_b = -\frac{1}{2\pi a_b h_b C_m} \begin{pmatrix} 0 \\ \vdots \\ 0 \\ g_{b1}(t)(\bar{v}_b(x_{b1}) - E_{b1}) \\ 0 \\ \vdots \\ 0 \\ g_{b2}(t)(\bar{v}_b(x_{b2}) - E_{b2}) \\ 0 \end{pmatrix}$$

where  $h_b = \ell_b/\gamma_b$  is the step size on branch  $b$ . This term arises as  $1/h_b$  when the delta function is discretized. Gathering the branch and soma equations together we define

$$\mathbf{Q} = \begin{pmatrix} \mathbf{Q}_1 & & \\ & \ddots & \\ & & \mathbf{Q}_B \\ & & & \mathbf{Q}_\sigma \end{pmatrix} \in \mathbb{R}^{N \times N}, \quad \mathbf{z} = \begin{pmatrix} z_1 \\ \vdots \\ z_B \\ z_\sigma \end{pmatrix} \in \mathbb{R}^N,$$

$$\mathbf{B} = \begin{pmatrix} \mathbf{B}_1 & & \\ & \ddots & \\ & & \mathbf{B}_B \\ & & & \mathbf{B}_\sigma \end{pmatrix} \in \mathbb{R}^{N \times n}, \quad \mathbf{u} = \begin{pmatrix} u_1 \\ \vdots \\ u_B \\ u_\sigma \end{pmatrix} \in \mathbb{R}^n.$$

and now coupling the branch conditions via the Hines matrix by  $\mathbf{A} = \mathbf{Q} + \mathbf{H}$  we arrive at the compartmental model

$$\mathbf{z}'(t) = \mathbf{A}\mathbf{z}(t) + \mathbf{B}\mathbf{u}(t). \quad (14)$$

We view  $\mathbf{u}$  as the input to the neuron. The neuron's output,  $\mathbf{y}$ , is the voltage at the site of initiation of action potentials, which is expressed as

$$\mathbf{y}(t) = \mathbf{C}\mathbf{z}(t). \quad (15)$$

We suppose that the output is the soma potential, formally written as  $\mathbf{y}(t) = \mathbf{z}_{N-m}(t)$ . This implies  $\mathbf{C} \in \mathbb{R}^{1 \times N}$  where  $\mathbf{C} = 0$  except for the entry  $\mathbf{C}_{1,N-m} = 1$ . Together Eqs. (14) and (15) give the standard form for linear dynamical systems, where  $\mathbf{z}$  is called the *state* vector,  $\mathbf{u}$  is the *input* vector, and  $\mathbf{y}$  is the *observable* vector. Our

goal is to reduce the dimension of this system while exactly preserving the inputs and without sacrificing accuracy of the soma potential.

### 3 Model reduction of the quasi-active system

We begin with a classical time-domain method known as Balanced Truncation (BT) and then contrast this with a newer frequency-domain approach deemed the Iterative Rational Krylov Algorithm.

#### 3.1 Balanced truncation

To **balance** a linear dynamical system is to transform it into one where the controllability and observability gramians coincide. As a result, and in a rigorous quantitative sense, the states that are difficult to reach are rarely observed. We present the method below. For the early history, and further details and applications, see (Kailath 1980) and (Antoulas and Sorensen 2001).

The controllability gramian  $\mathcal{P}$  and observability gramian  $\mathcal{Q}$  are defined via

$$\mathcal{P} = \int_0^\infty e^{\mathbf{A}t} \mathbf{B} \mathbf{B}^* e^{\mathbf{A}^* t} dt, \quad \mathcal{Q} = \int_0^\infty e^{\mathbf{A}^* t} \mathbf{C}^* \mathbf{C} e^{\mathbf{A}t} dt,$$

but typically are computed as solutions to the Lyapunov equations

$$\mathbf{A}\mathcal{P} + \mathcal{P}\mathbf{A}^* + \mathbf{B}\mathbf{B}^* = 0, \quad \mathbf{A}^*\mathcal{Q} + \mathcal{Q}\mathbf{A} + \mathbf{C}^*\mathbf{C} = 0.$$

We gather their Cholesky factors

$$\mathcal{P} = \mathbf{U}\mathbf{U}^*, \quad \text{and} \quad \mathcal{Q} = \mathbf{L}\mathbf{L}^*.$$

and compute the singular value decomposition of the mixed product

$$\mathbf{U}^* \mathbf{L} = \mathbf{Z} \mathbf{\Sigma} \mathbf{Y}^*.$$

Here  $\mathbf{\Sigma}$  is a diagonal matrix whose entries are the eigenvalues of  $\mathbf{U}^* \mathcal{Q} \mathbf{U}$ ,  $\mathbf{Z}$  is an orthogonal matrix whose columns are the eigenvectors of  $\mathbf{U}^* \mathcal{Q} \mathbf{U}$ , and  $\mathbf{Y}$  is an orthogonal matrix whose columns are the eigenvectors of  $\mathbf{L}^* \mathcal{P} \mathbf{L}$  (for details on the SVD, see (Trefethen and Bau 1997)). The diagonal elements of  $\mathbf{\Sigma}$ , are nonnegative and in descending order and are known as the **Hankel Singular Values** (HSV's) of the dynamical system described by Eqs. (14) and (15). We now compose

$$\mathbf{T} = \mathbf{\Sigma}^{-1/2} \mathbf{Y}^* \mathbf{L}^*, \quad \text{and} \quad \mathbf{T}^{-1} = \mathbf{U} \mathbf{Z} \mathbf{\Sigma}^{-1/2},$$

and note that the transformed gramians

$$\hat{\mathcal{P}} = \mathbf{T} \mathcal{P} \mathbf{T}^* \quad \text{and} \quad \hat{\mathcal{Q}} = \mathbf{T}^{-*} \mathcal{Q} \mathbf{T}^{-1}$$

are balanced and diagonal in the sense that

$$\widehat{\mathcal{P}} = \widehat{\mathcal{Q}} = \Sigma.$$

Moreover they are the gramians of the transformed state,  $\widehat{\mathbf{z}} = T\mathbf{z}$ , which itself is governed by the transformed dynamical system as

$$\frac{d\widehat{\mathbf{z}}(t)}{dt} = \widehat{\mathbf{A}}\widehat{\mathbf{z}}(t) + \widehat{\mathbf{B}}\mathbf{u}(t), \quad \mathbf{y}(t) = \widehat{\mathbf{C}}\widehat{\mathbf{z}}(t) \quad (16)$$

where  $\widehat{\mathbf{A}} = T\mathbf{A}T^{-1}$ ,  $\widehat{\mathbf{B}} = T\mathbf{B}$ ,  $\widehat{\mathbf{C}} = \mathbf{C}T^{-1}$ . Based on the decay of the singular values in  $\Sigma$ , we can construct a reduced model by using only the  $k$  largest singular values. This corresponds to approximating Eq. (16) with

$$\frac{d\widehat{\boldsymbol{\xi}}(t)}{dt} = \widehat{\mathbf{A}}_{11}\widehat{\boldsymbol{\xi}}(t) + \widehat{\mathbf{B}}_1\mathbf{u}(t) \quad \widehat{\mathbf{y}}(t) = \widehat{\mathbf{C}}_1\widehat{\boldsymbol{\xi}}(t), \quad (17)$$

where  $\widehat{\mathbf{A}}_{11}$  is the initial  $k \times k$  submatrix of  $\widehat{\mathbf{A}}$ ,  $\widehat{\mathbf{B}}_1$  is the first  $k$  rows of  $\widehat{\mathbf{B}}$ , and  $\widehat{\mathbf{C}}_1$  is the first  $k$  columns of  $\widehat{\mathbf{C}}$ .

The reduced state vector  $\boldsymbol{\xi} \in \mathbb{R}^k$  has no apparent physiological sense until it is fed to the equation for  $\widehat{\mathbf{y}}$ , but when  $k \ll n$  the computational savings from this model reduction is great. Moreover, contrary to the brute force approach of compartment coarsening, we have not sacrificed the rich input structure. In Section 4.1 we demonstrate that the system in Eq. (17) allows for numerically exact (near machine precision) reproduction of the soma potential computed from the quasi-active system in a fraction of the time.

### 3.2 The iterative rational Krylov algorithm

For small systems, Balanced Truncation is a clean, precise, and feasible method, but as the system dimension grows the cost of computing the gramians becomes prohibitive. Since the reduction comes after the gramians have been computed, this requires storage of dense  $N \times N$  matrices before the reduction step, meaning that memory is also an issue.

Krylov methods, on the other hand, construct approximate reduced systems of a given dimension from the beginning. Instead of transforming the original system and then truncating, the Krylov process iteratively projects the dynamics of the original system onto a smaller subspace. This reduces the memory requirements significantly, since only matrix-vector products are used, and in turn drastically speeds up the reduction process. We now describe the main ideas behind this algorithm; for full details see (Gugercin et al. 2008).

Consider the quasi-active system given in Eqs. (14) and (15). We construct two matrices,  $\mathbf{V}_k, \mathbf{W}_k \in \mathbb{R}^{N \times k}$  such that  $\mathbf{z}(t) = \mathbf{V}_k\boldsymbol{\xi}(t)$  for some  $\boldsymbol{\xi}(t) \in \mathbb{R}^k$  and such that

$$\mathbf{W}_k^T(\mathbf{V}_k\boldsymbol{\xi}'(t) - \mathbf{A}\mathbf{V}_k\boldsymbol{\xi}(t) - \mathbf{B}\mathbf{u}(t)) = 0 \quad (18)$$

and

$$\text{Range}(\mathbf{V}_k) \cap \text{Null}(\mathbf{W}_k^T) = \{0\}. \quad (19)$$

From Eq. (19) we see that  $\mathbf{W}_k^T\mathbf{V}_k$  is invertible. Hence we can use Eq. (18) to construct a reduced model

$$\frac{d\boldsymbol{\xi}(t)}{dt} = \mathbf{A}_k\boldsymbol{\xi}(t) + \mathbf{B}_k\mathbf{u}(t), \quad \widehat{\mathbf{y}}(t) = \mathbf{C}_k\boldsymbol{\xi}(t). \quad (20)$$

where

$$\mathbf{A}_k = (\mathbf{W}_k^T\mathbf{V}_k)^{-1}\mathbf{W}_k^T\mathbf{A}\mathbf{V}_k, \quad \mathbf{B}_k = (\mathbf{W}_k^T\mathbf{V}_k)^{-1}\mathbf{W}_k^T\mathbf{B},$$

$$\mathbf{C}_k = \mathbf{C}\mathbf{V}_k. \quad (21)$$

The reduced-order system is computed by finding  $\mathbf{V}_k$  and  $\mathbf{W}_k$  so that the  $L_2$ -norm of the error between the transfer functions of the original and reduced systems along the imaginary axis is minimized, i.e. we solve the optimization problem

$$\min_{\mathbf{V}_k, \mathbf{W}_k} \int_{-\infty}^{\infty} \|\mathbf{C}^T(i\omega\mathbf{I} - \mathbf{A})^{-1}\mathbf{B} - \mathbf{C}_k^T(i\omega\mathbf{I} - \mathbf{A}_k)^{-1}\mathbf{B}_k\|^2 d\omega.$$

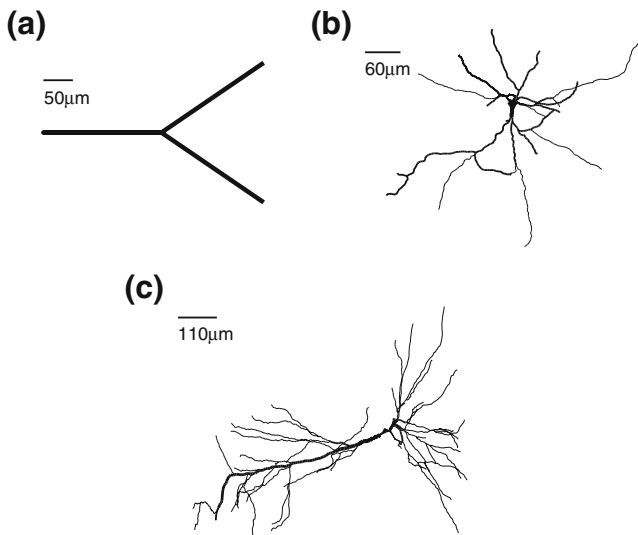
One strategy for solving this is to interpolate the full transfer function, to first order, at the negative of each of its poles. Since these poles are not generally known *a priori* and may be hard to compute, we make an initial guess and then iterate until convergence, indicating that we have arrived at the reduced system. This is achieved in a computationally efficient manner via the Iterative Rational Krylov Algorithm (IRKA), whose implementation details we do not discuss here but are found in (Gugercin et al. 2008).

## 4 Balanced truncation model reduction results

We have written a MATLAB suite of functions that loads morphology and channel kinetics and distributions, constructs the associated quasi-active system and both its BT and IRKA reductions, and computes and displays the response of the 4 (nonlinear, quasi-active, BT, IRKA) models to random (in space and time) sequences of synaptic input. These codes are available from the authors upon request. All computations were performed on a Sun Ultra 20 computer with a 2.2 GHz AMD Opteron processor.

Morphologies, shown in Fig. 2, were obtained from the Rice-Baylor archive (Martinez 2008) and from NeuroMorpho.org (<http://NeuroMorpho.org>; Ascoli 2006), and then imported to our software suite, which lets the user visualize and simulate the neuron via GUI's. Pyramidal and interneuron ion channel models were obtained from the literature and from ModelDB (<http://senselab.med.yale.edu/modeldb>; Hines et al.





**Fig. 2** Renderings of the three morphologies used in this paper: (a) a forked neuron; (b) cell AR-1-20-04-A, a projection neuron from rat entorhinal cortex (Martinez 2008); (c) cell n408, a rat hippocampal cell from region CA1 (Pyapali et al. 1998)

2004), as detailed in Tables 2 and 3 of the Appendix. Unless otherwise indicated,  $\bar{g}_{bs} = 0$ .

#### 4.1 Dimension reduction ratio

Consider a forked neuron as shown in Fig. 2(a). Each of the three branches is 200  $\mu\text{m}$  long and is divided into 2  $\mu\text{m}$ -long segments. The root has radius 2  $\mu\text{m}$  while the leaves have radius 1  $\mu\text{m}$ . Ion channels with Hodgkin–Huxley kinetics (see Table 2) were uniformly distributed. This leads to a quasi-active system of dimension 1204. We computed BT matrices and found that the Hankel singular values decay quite rapidly (Fig. 3(a)). Indeed,  $\sigma_n < \epsilon_{\text{mach}}$  for  $n > 65$ , indicating that a reduction of at least 20-fold can be obtained. Numerically we observe in Fig. 3(b) that, compared to the computed quasi-active soma potential, nearly 5 digits of accuracy can be obtained using only 12 HSV's, a reduction of fully two orders of magnitude. The BT dynamics also track the dynamics of the soma potential from the nonlinear system qualitatively well, though the error is larger (Fig. 3(c)).

Tonic synapses have a significant effect on the dynamics, but no effect on the accuracy of the reduced model versus the quasi-active one. We demonstrate the impact of tonic synapses by randomly choosing 10% of the compartments to have tonic components  $\bar{g}_{bs} = 0.2$  nS, which changes the rest potential of the neuron from a uniform value of  $\approx -64.9186$  mV to one that is now spatially-varying and elevated (Fig. 3(d)).

Using the exact same stimulus as used for Fig. 3(b–c), we find that the neuron's output is now more oscillatory, but the reduced system's accuracy is unchanged (Fig. 3(e–f)).

Computing the BT matrices required about 72 s, while the 30 ms simulation required only 0.02 s. At first glance this appears expensive when compared to the nonlinear and quasi-active simulation times, both requiring about 1.3 s. However, the BT matrix computation is a one-time cost, since these matrices can now be reused to facilitate simulation with their associated morphology.

In fact, it can be seen that the decay of the normalized Hankel singular values almost directly corresponds with the numerical accuracy achieved, as Fig. 4 illustrates on a more realistic neuron. This result is not surprising, however, given that an error bound exists for BT model reduction (Antoulas and Sorensen 2001). Therefore, the normalized Hankel singular values may be used as a reliable guide to achieving any desired numerical accuracy compared to the quasi-active system.

#### 4.2 Application to synaptic scaling

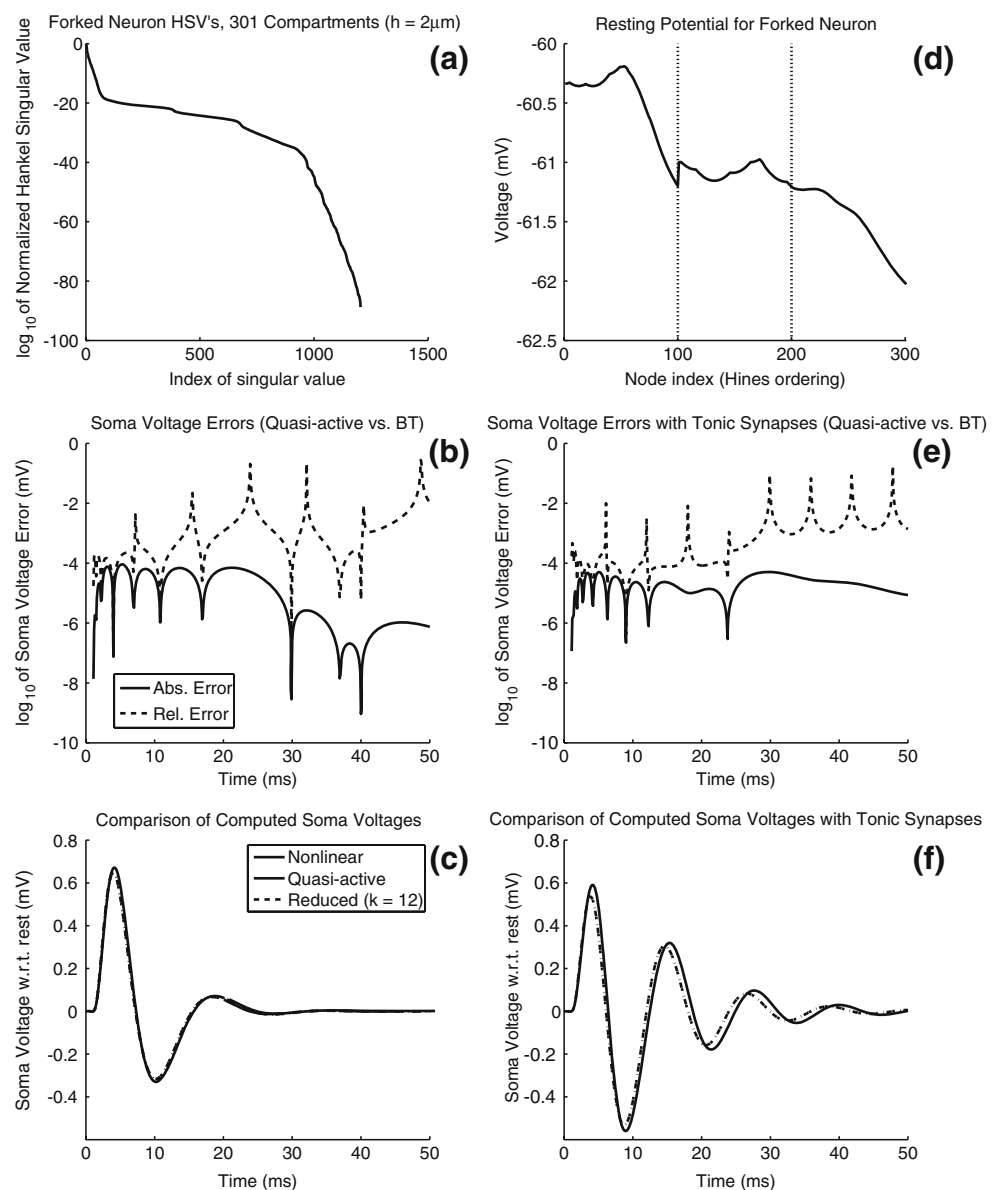
An immediate application for such low-dimensional systems is to accurately quantify how synaptic input scales with distance to the soma, also known as “dendritic democratization” (Magee and Cook 2000; Häusser 2001; Timofeeva et al. 2008). One standard form of synaptic input is an alpha function, which describes the input onto synapse  $s$  of branch  $b$  as

$$\tilde{g}_{bs}(t) = \hat{g}_{bs} \frac{t - t_{\text{start}}}{\tau} \exp\left(1 - \frac{t - t_{\text{start}}}{\tau}\right), \quad (22)$$

where  $t_{\text{start}}$  is the time (in ms) of stimulus onset, and the time constant  $\tau = 1$  ms is the time at which the conductance reaches its maximum value of  $\hat{g}_{bs}$ . In order to know the strength required for a single synaptic input at a given location to produce a peak target depolarization at the soma, we need to run a parameter sweep for  $\hat{g}_{bs}$ . Using the reduced system on the forked neuron with the uniform channels of Table 2, we execute the bisection method with a tolerance of  $10^{-8}$  to obtain values for each compartment, shown in Fig. 5, in a total of 4.7 min. By contrast, the nonlinear system would require about 100 min to do the same.

For more realistic morphologies, and more complex channel models, the one-time cost of computing the BT matrices grows as  $\mathcal{O}(n^3)$ , but the speed-up gained from using the resulting reduced system is substantial. As

**Fig. 3** (a) Hankel singular values for the forked neuron given a  $2\ \mu\text{m}$  compartment size. Since there are 301 compartments, and the HH ion channels have 3 gating variables, the quasi-active system has 1204 variables, hence the BT system has the same number of HSV's. (b) Absolute and relative errors between the computed soma potentials of the quasi-active and BT systems. (c) Computed soma potentials from the nonlinear, quasi-active, and BT simulations. At the 1 ms mark a 1 nS alpha-function synaptic input with time constant  $\tau = 1\ \text{ms}$  was applied to the midpoint of the top distal branch. Notice that the quasi-active and BT curves are indistinguishable. (d) Resting potential of each node for the forked neuron with 10% tonic synapses of  $\bar{g}_{bs} = 0.2\ \text{nS}$  each. In the absence of tonic synapses, each node is at rest at  $\approx -64.9186\ \text{mV}$ . The faint dotted line indicates the end of a branch. (e–f) Same as (b–c), but for the case of tonic synapses. Notice that tonic synapses induce more oscillations for the exact same stimulus

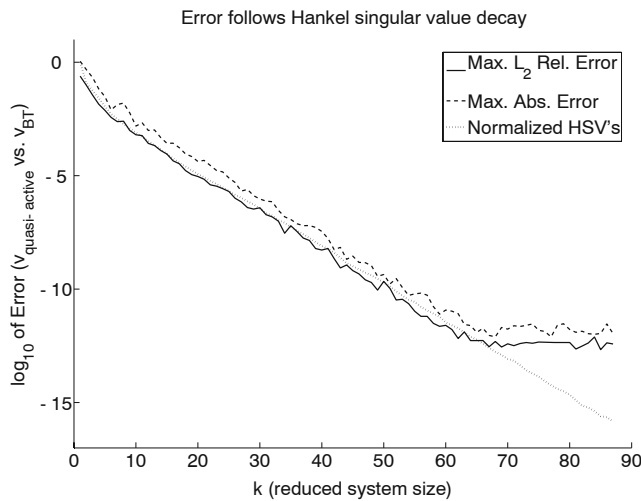


an example, neuron AR-1-20-04-A (see Fig. 2(b)) was discretized into 1121 compartments, each with length  $\approx 2\ \mu\text{m}$ . We included  $I_{Na}$  and  $I_K$  currents following Hodgkin–Huxley kinetics, and used Connor–Stevens kinetics (see Table 3) (Connor and Stevens 1971) for the A-type  $K^+$  current  $I_A$  whose spatial distribution followed that suggested by (Hoffman et al. 1997). The resulting system had 5 gating variables per compartment, and hence the dimension of the linearized system was 6726. The BT matrices were computed in about 4.6 hours, and setting  $k = 25$  we obtained a reduced system that was  $269\times$  smaller than the original linearized system and accurate to more than 5 digits.

Using this reduced model, we executed a parameter sweep for  $\hat{g}_{bs}$  for each compartment, taking a total of

42 min. To compare the output to that of the nonlinear model, we did a sweep for 7 compartments in each branch, taking 268 total minutes. The results indicate (see Fig. 6) that the reduced system accurately tracked the scaling of proximal synaptic inputs while underestimating the scaling for distal conductances. The error in the distal computations is not due to our reduction scheme but rather to our initial quasi-active hypothesis. Regarding Fig. 6, a value of  $\hat{g}_{bs}$  for every compartment was obtained using the reduced model, whereas to do so for the active model would have required about 21 hours. Hence, to obtain a full portrait of synaptic scaling for this neuron, even factoring in the cost of computing the BT matrices, the reduced system offers a factor of 4 speed-up.

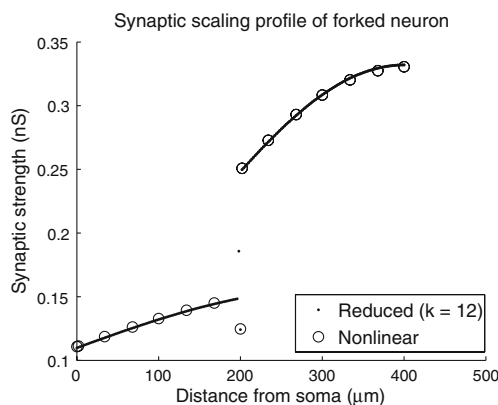




**Fig. 4** Simulations demonstrate converging accuracy for the quasi-active versus reduced system. We used the neuron of Fig. 2(b) with the nonuniform channel model of Table 3. The discretized neuron had 1121 compartments, leading to a 6726-dimensional system. For each reduced system size  $k$ , we performed 10 simulations of 50 ms each with 35 random alpha-function synaptic inputs for which  $\hat{g}_{bs} \in [0, 2]$

#### 4.3 Application to dendritic resonance

Neurons have been shown to respond preferentially to stimuli occurring at specific frequencies (Hutcheon and Yarom 2000; Hasselmo et al. 2007; Ulrich 2002), and quantifying how this resonant frequency varies with stimulus location is a difficult experimental task. However, our reduced model is well-suited to this type of simulation.



**Fig. 5** Strengths  $\hat{g}_{bs}$  of alpha-function synaptic input needed at each compartment to get a 0.2 mV peak soma depolarization. Dots were computed using the BT system, while circles were computed using the nonlinear system. Notice that the conductance on the root branch increases slowly with increasing distance from the soma, but at the branch point this trend splits. The large difference at the junction occurs because of the disjoint radii of the root branch and the child branches. Only one distinct set of points can be seen from 200–400  $\mu\text{m}$  because the child branches are identical

Acquiring the resonant frequency data is time-consuming if high frequency resolution is desired. Typical experimental studies stimulate the cell with a ZAP current, which has the form

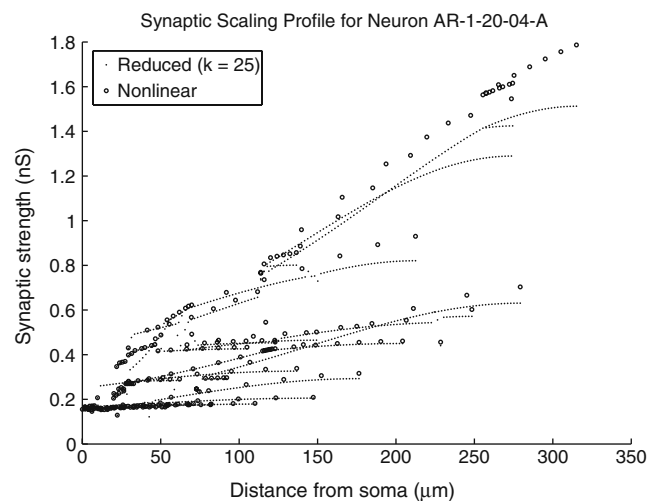
$$I(t) = a \sin(bt^c + d) \quad (23)$$

for some constants  $a, b, c$ , and  $d$ , with  $t$  in ms. The resonant frequency can then be obtained by dividing the Fourier transform of the voltage by the Fourier transform of the input current (Puil et al. 1986). However, this limits the frequency resolution unless very long or detailed simulations are performed. On the other hand, brute force methods give better resolution, but are slow because require many simulations to be performed using oscillatory currents, such as

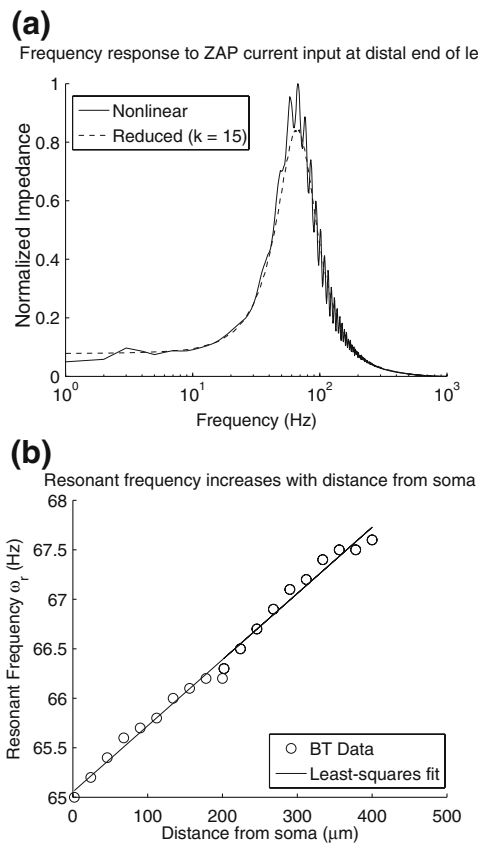
$$I(t) = I_0(0.5 + 0.5 \sin(2\pi\omega(t - \hat{t})/1000)), \quad (24)$$

where  $I_0$  is the peak amplitude,  $\hat{t}$  is the time of stimulus onset (in ms), and  $\omega$  is the frequency (in Hz).

Our reduced model is useful here because we can use the ZAP current to get a band in which the resonant frequency  $\omega_r$  may be and then run a brute-force search to pinpoint this value, all in much less time than the full system requires. For example, using the ZAP current in Eq. (23) with parameters  $a = 50$ ,  $b = 10^{-6}$ , and  $c = 3$ , we find that the forked neuron with uniform channels has a definite peak near 65 Hz (see Fig. 7(a)). Refining this estimate via a brute-force search using Eq. (24) in a small interval around this peak reveals that  $\omega_r$  increases linearly as distance from the soma increases.

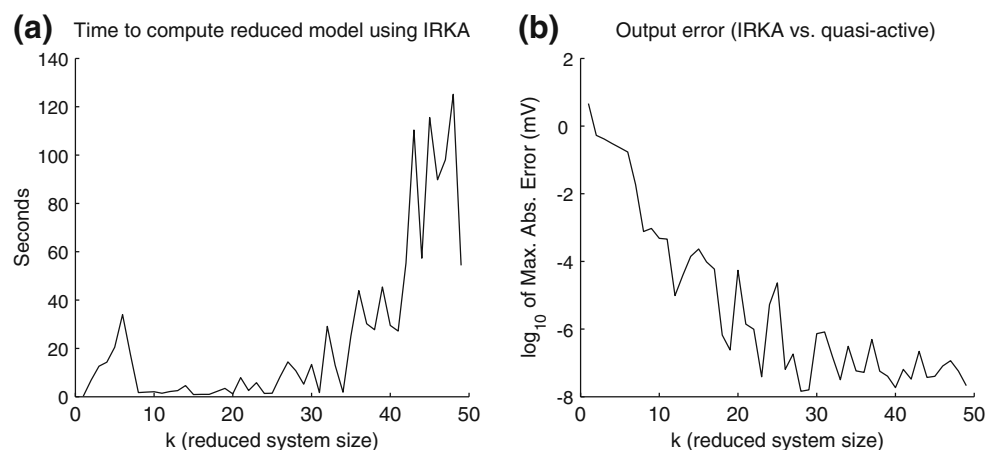


**Fig. 6** Strengths  $\hat{g}_{bs}$  of alpha-function synaptic input needed at each compartment of neuron AR-1-20-04-A to get a 0.2 mV soma depolarization using the BT method. Dots were computed using the reduced system, while circles were computed using the nonlinear system. The reduced system size was  $k = 25$



**Fig. 7** (a) Frequency-response curves for ZAP current injection at different dendritic locations on the forked neuron with uniform channels. We applied ZAP current for 1000 ms with timestep  $\Delta t = 0.1$  ms and computed the normalized impedance versus frequency. Both curves show the same trend, with a peak near 65 Hz, but the reduced system took only 2 s, whereas the nonlinear system took 39 s. (b) Scatter plot of resonant frequencies versus distance from the soma (circles), with the least-squares linear fit to the data (thick line). For this experiment, we sampled frequencies every 0.1 Hz on the interval [55, 75] using the current in Eq. (24) and found that the resonant frequency increases nearly linearly as the stimulus location moves away from the soma

**Fig. 8** (a) Time to compute reduced system, and (b) maximum absolute error between quasi-active and reduced system output using IRKA for cell AR-1-20-04-A ( $N = 6726$ )

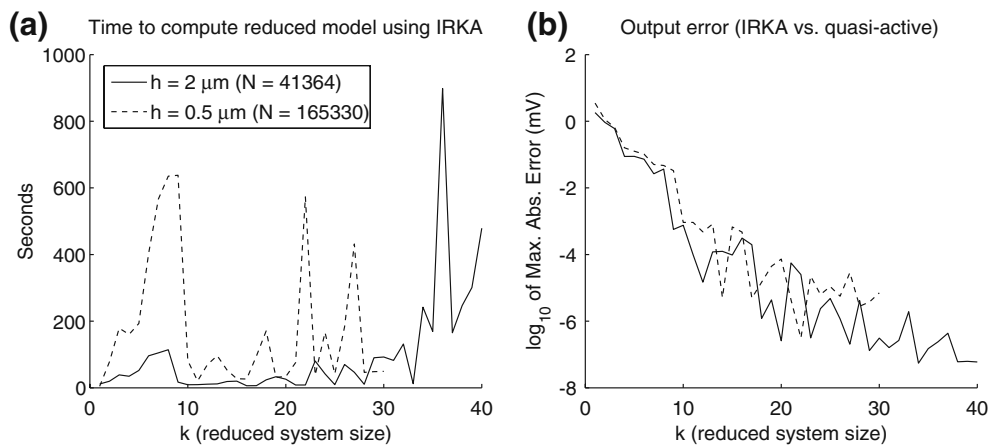


## 5 IRKA model reduction results

With the BT method as a benchmark, we test the model reduction technique based on using IRKA. First we demonstrate typical convergence of IRKA for our problems, and then we show that cells of much larger dimension can be tackled.

Consider neuron AR-1-20-04-A from Fig. 2(b). We compute the maximum absolute error in the soma voltage between the quasi-active and reduced systems using IRKA (Fig. 8(b)), and we compare this to the error curve shown in Fig. 4(a). Notice that nearly 5-digit agreement occurs for a 25-dimensional system using IRKA. This is close to what we would expect from the BT method (20-dimensional system), but IRKA required only 5 s to compute the reduced system (Fig. 8(a)), whereas BT required 4.6 h. Thus similar accuracy with IRKA can be obtained in a fraction of the time needed for BT. Notice that although the error in IRKA as  $k$  increases is not as smooth of a trend as that of BT, nor does the error decrease as rapidly, the accuracy is more than enough for neuroscience applications.

The main difference between IRKA and BT is that IRKA computes a system of a given size, whereas BT computes full matrices that are then truncated to the desired size. If a reduced system of a different size is desired, IRKA must recompute the matrices, whereas BT can just truncate. But the premium we pay for immediate BT changes is far too expensive, considering that IRKA's results are just as good but are obtained in a fraction of the time. As an example, the reduced system from IRKA with  $k = 20$  was used to run the same dendritic democratization parameter sweep as in Fig. 6. The computed synaptic conductances for the BT and IRKA systems agree to nearly 5 digits for each



**Fig. 9** (a) Time to compute the reduced system, and (b) maximum absolute error between the quasi-active and reduced system output using IRKA on cell n408. Solid lines correspond to the  $N = 41364$  case, while dashed lines are for the  $N = 165330$  case.

compartment, indicating that IRKA is indeed computing the right reduced system.

A clear advantage of IRKA is that it handles systems of much larger dimension than BT. We could not use BT for systems where  $N > 7000$ , due to lack of memory, and even if we could, the computation time would preclude any practical use for such large systems. IRKA does not suffer from this drawback, which translates into the ability to compute reduced models of cells with much finer discretizations and to reduce cells with much more complex dendritic structure.

As an example, we consider neuron n408 (Fig. 2(c)), a rat hippocampal cell from region CA1 (Pyapali et al. 1998). We use a  $2 \mu\text{m}$  spatial discretization, yielding 6894 compartments. With a Connor-Stevens ion channel model this gives a total linear system size of  $N = 41364$ . Using IRKA we find that 5 digits of accuracy can be obtained using a system of dimension  $k \approx 15$ , which is a 2700-fold reduction, or more than 3 orders of magnitude, in less than a minute (see Fig. 9). If we use a finer discretization of  $h \approx 0.5 \mu\text{m}$ , we arrive at a system of size  $N = 165330$ . IRKA produces a 15-dimensional system that is accurate to nearly 5 digits, a monstrous 11000-fold reduction, or 4 orders of magnitude!

## 6 A quasi-integrate-and-fire model

Quasi-active neuron models cannot spike, but adopting an integrate-and-fire (IAF) mechanism allows such models to emulate this behavior. Under our assumption

In (a) the sudden drop in computation time around  $k = 10$  occurs for both cases because at that point, though the shifts have not converged, the reduced system is accurate to a tolerance of 3 digits or more

that the soma is the site of action potential generation, we only need to check whether

$$\hat{\mathbf{y}}(t) = \tilde{v}_\sigma(t) \geq V_{\text{th}} \quad (25)$$

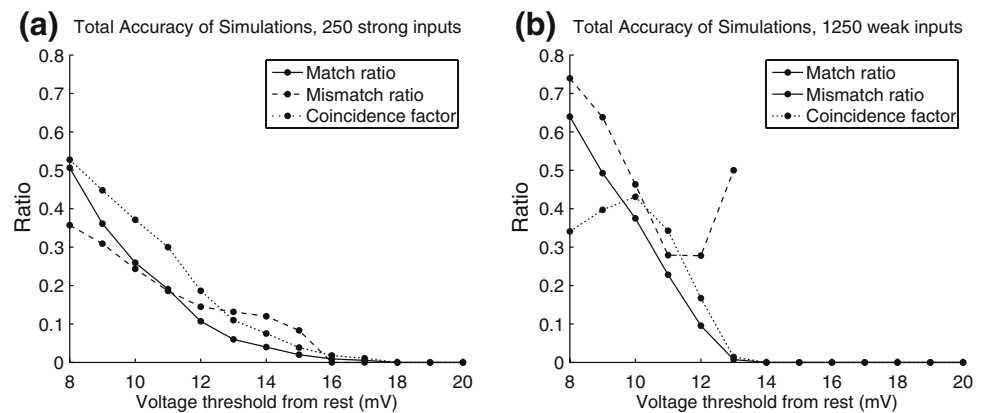
where  $V_{\text{th}}$  is some threshold voltage (recall that  $\tilde{v}_\sigma(t) \equiv \tilde{v}_1(\ell_{b_1^1}, t)$ ). At each timestep during the simulation we check if threshold was reached, and if so then we instantaneously reset the soma voltage to the rest and hold it there for a refractory period of  $\tau_{\text{ref}}$  ms.

### 6.1 Thresholding at the soma

Reset values should be chosen to produce as much similarity as possible between the outputs of the active and quasi-active models while still being biophysically reasonable. However, the state variables of the reduced system do not have biophysical significance until transformed into the observable  $\hat{\mathbf{y}}(t)$ . This implies our choice of rest as the reset value for  $\xi(t)$ .

Using the forked neuron as our test case, we implement the IAF mechanism using the non-uniform channel model. We discretized using  $h \approx 1 \mu\text{m}$  so that the nonlinear model had 3606 variables; IRKA computed a reduced model having only 10 variables. First we simulate using the nonlinear system, and then we compare this to the output from the reduced system simulation. The spikes generated from both systems are compared to determine if there are any matches. A match is said to occur when the reduced system's spike occurs within  $\tau_{\text{ref}}$  ms before or after a spike in the nonlinear system.

**Fig. 10** Ratio of matched and mismatched spikes, as well as coincidence factor, as  $V_{th}$  increases for (a) 250 “strong” inputs and (b) 1250 “weak” inputs per simulation. For each  $V_{th}$  value we used  $\tau_{ref} = 4$  ms and ran at least 20 simulations of 1000 ms with alpha-function synaptic inputs to random locations. Ratios are given for the match and mismatch values rather than percentages for ease of comparison with  $\Gamma$



In order to quantify how well the mechanism captures spiking behavior, we compute the coincidence factor  $\Gamma$  (Kistler et al. 1997; Jolivet et al. 2006), defined as

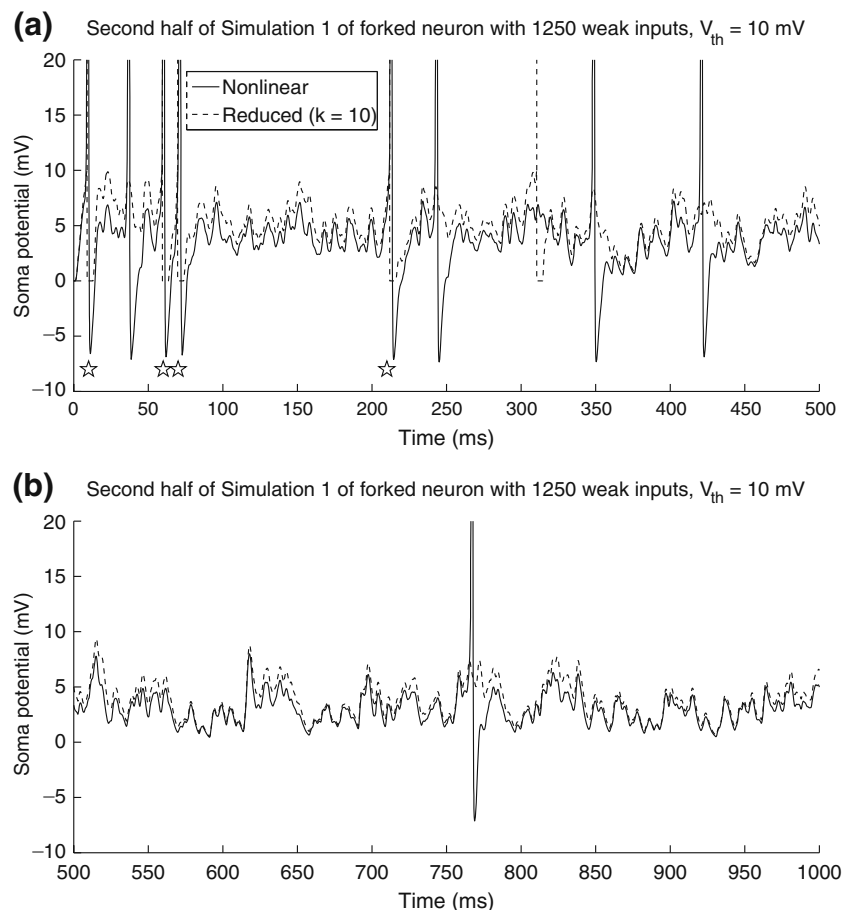
$$\Gamma = \frac{N_{\text{coinc}} - N_{\text{nonlin}} N_{\text{reduced}} (\tau_{\text{ref}}/T)}{(N_{\text{nonlin}} + N_{\text{reduced}})(1 - N_{\text{nonlin}} (\tau_{\text{ref}}/T))/2}, \quad (26)$$

where  $N_{\text{nonlin}}$  and  $N_{\text{reduced}}$  are the number of spikes in the nonlinear and reduced models, respectively, and  $T$  is the length of the simulation. The coincidence factor

measures how close the spike train from the reduced model approximates that of the nonlinear model by comparing the number of coincident spikes,  $N_{\text{coinc}}$ , with the number of coincident spikes occurring by chance.  $\Gamma$  is scaled to ensure that  $\Gamma = 1$  implies the spike trains are equal and  $\Gamma = 0$  implies the spike trains would occur purely by chance.

However,  $\Gamma$  can be quite low even when most spikes from the reduced system are coincident. Thus, to better evaluate the effect of false positives, we introduce two

**Fig. 11** (a–b) Soma potentials corresponding to the first and second halves of simulation 1 of the “weak” input case using  $V_{th} = 10$  mV. Note that the reduced system spikes are drawn for clarification, and that the voltage axis has been scaled so that subthreshold behavior is more easily seen. Subthreshold behavior of the nonlinear model is captured well by the reduced model, especially when less firing occurs, and stars below spikes in (a) indicate that four spikes have been captured by the reduced model



metrics: percentage matched, which is the number of spikes from the nonlinear system that are also found in the reduced system; and percentage mismatched, which is how many spikes from the reduced system do not match any spikes from the nonlinear system:

$$\% \text{ matched} = 100 \frac{\# \text{ matched}}{\# \text{ nonlinear spikes}} \quad (27)$$

$$\% \text{ mismatched} = 100 \frac{\# \text{ unmatched reduced spikes}}{\# \text{ reduced spikes}}. \quad (28)$$

We vary  $V_{th}$  between 8 and 20 mV, with at least 20 simulations at each threshold value, and we use  $\tau_{ref} = 4$  ms throughout. Each simulation lasts 1000 ms, during which alpha-function synaptic inputs arrive at random locations and at random times.

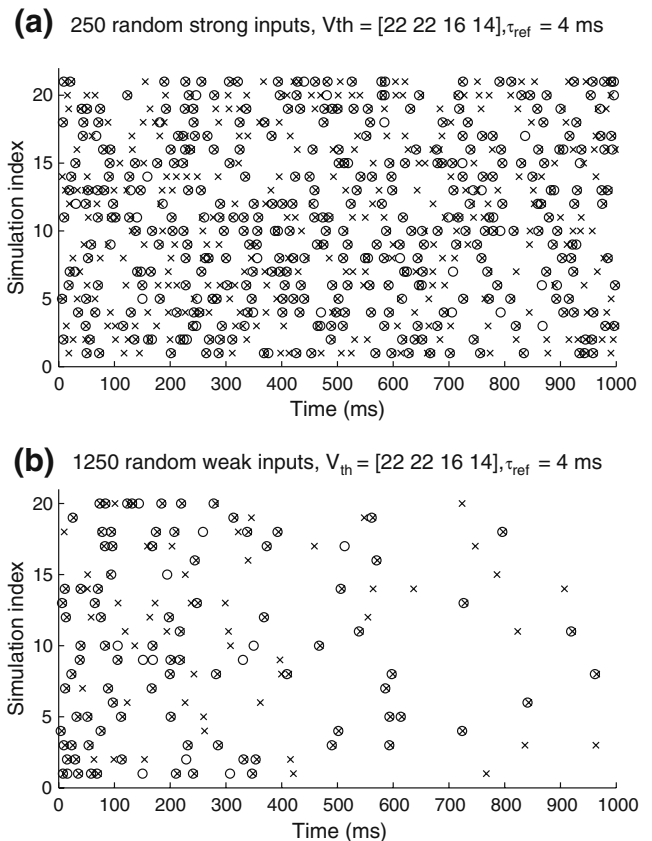
We ran two sets of simulations to assess the effect of low- and high-activity inputs. The low-activity set used 250 “strong” inputs per simulation, while the high-activity one used 1250 “weak” inputs. These inputs are realistically calibrated by computing  $\hat{g}_{bs}$  values for the reduced system (as in Section 4.2). These values are scaled to obtain synaptic conductances at each compartment that would give approximately 3 mV depolarizations at the soma for the strong inputs, but only 1 mV depolarizations for the weak ones, which agrees with typical values in Chapter 3 of (Traub and Miles 1991) and (Thomson and Lamy 2007).

As shown in Fig. 10, low thresholds capture a large number of the nonlinear spikes, but also yield more mismatched spikes. Fewer mismatched spikes occur as  $V_{th}$  increases, but this also leads to greatly diminished numbers of spikes generated, and hence fewer nonlinear spikes captured. More mismatched spikes were observed in the “weak” input simulations, but more spikes were matched as well. The coincidence factors were also low for these simulations, reaching a peak of 0.43 at  $V_{th} = 10$  mV for the “weak” input case and a peak of 0.52 at  $V_{th} = 8$  mV for the “strong” input case.

The spike data is shown explicitly in the soma potentials of Fig. 11. It is evident from these plots that the reduced system captures the subthreshold behaviors very well even for large depolarizations. Furthermore, the spike times computed by the reduced system tend to be very close to the actual spike times.

## 6.2 Thresholding at multiple sites

Of course one has no right to expect high accuracy when applying supra-threshold stimuli to a sub-threshold model. Nonetheless, there are a number of means by which our crude mechanism may be improved. If we fail to detect a somatic spike because



**Fig. 12** Raster diagrams showing spikes computed for the nonlinear system (crosses) and the reduced system (circles) for (a) “strong” and (b) “weak” input cases demonstrate that a variable-threshold, multi-site IAF mechanism improves accuracy of spike generation. The same input patterns as used for Fig. 10 were applied to the forked neuron, but this time voltages were observed at the soma and the midpoints of each branch. Threshold values are given in Section 6.2. In total, the “strong” and “weak” input simulations had match scores of 56% and 65%, respectively, and mismatch scores of 15% and 13%, respectively

it was generated in the dendrite, then it may pay to build the reduction in order that it accurately tracks the quasi-active response at several, say  $p$ , distinct locations.

Proper investigation of this issue would address the questions of optimal placement of  $p$  observers, placement and calibration of threshold mechanisms at each site, and the trade of accuracy for speed as  $p$  is increased. We offer here only empirical evidence that such an investigation is indeed warranted.

In particular, we apply a multi-site IAF mechanism to the forked neuron subject to the same input streams

**Table 1** A multi-site variable-threshold IAF mechanism improves spike-capturing accuracy

Input scheme	$N_{nonlin}$	$N_{reduced}$	% Matched	% Mismatched	$\Gamma$
1250 weak	136	101	64.7	12.9	0.73
250 strong	582	385	56.2	15.1	0.66



that generated Fig. 10. Since there are 3 branches and a soma, we set  $p = 4$  and designate as observables the voltages at the soma and the midpoints of each branch. Threshold values are  $V_{th} = 22$  mV at the leaf midpoints,  $V_{th} = 16$  mV at the root midpoint, and  $V_{th} = 14$  mV at the soma (these values were chosen after manual trial-and-error). Using the same discretization as in Section 6.1, IRKA computes a reduced model with  $k = 40$  that is accurate to nearly 5 digits, and hence  $\mathbf{A}_k \in \mathbb{R}^{40 \times 40}$ ,  $\mathbf{B}_k \in \mathbb{R}^{40 \times 601}$ , and  $\mathbf{C}_k \in \mathbb{R}^{4 \times 40}$ .

Using the same input patterns as above, we find a significant increase in the accuracy of the spiking behavior (see Fig. 12) with very little change in computational cost. For the strong input case we match 56% of the actual spikes and have only a 15% mismatch rate. For the weak input case, we get 65% matched with 13% mismatched. The coincidence factor improves in both cases: from 0.43 to 0.73 for the “weak” input case, and from 0.52 to 0.66 for the “strong” input case. Moreover, the difference in simulation time was negligible when compared to the thresholding mechanism of Section 6.1 (Table 1).

Rather than simple thresholding we note that a number of investigations of single compartment models have achieved better spike capturing by incorporating more sophisticated firing mechanisms. For example, (Fourcaud-Trocmé et al. 2003) includes a voltage-dependent exponential term, while (Jolivet et al. 2006) and (Brette and Gerstner 2005) consider adaptive thresholding. For the class of morphologies and channel distributions of the previous sections, adaptive thresholding, in our hands, did not produce a significant improvement in spike accuracy. Regarding the implementation of exponential integrate-and-fire we note that there is not a natural means by which to nonlinearize our reduced quasi-active system. More precisely, as we observed at the close of Section 3.1, the reduced state  $\xi$  in Eq. (17) is governed by the small but dense pair  $\hat{\mathbf{A}}_{11}$  and  $\hat{\mathbf{B}}_1$  and hence reflects a complex combination of thousands of gating and voltage variables. The physiological soma potential does not surface until after multiplication by  $\hat{\mathbf{C}}_1$ . As such there is not a distinguished voltage term in the differential

equation for  $\xi$  to which one might apply a biophysically inspired firing mechanism.

## 7 Discussion

We have applied two distinct methods to the reduction of dimension of large scale single neuron models. We have demonstrated that in typical settings one may reduce models with tens of thousands of variables to models with merely tens of variables that still accurately track the somatic subthreshold response of spatially distributed synaptic inputs. In regimes where subthreshold stimuli dominate, such as questions of (proximal) dendritic democratization and resonance, our reduced model reveals in minutes what the full models require hours or days to simulate. Although there is no reason to trust our subthreshold integrator on suprathreshold stimuli, we have demonstrated that simply thresholding the reduced voltage at a number of observation points has the potential to capture the cell’s full spiking behavior. Further refinement of this approach holds the promise of bringing morphological specificity to the large scale network contributions of (Rudolph and Destexhe 2006) with significantly fewer compartments than the approach proposed by (Markram 2006).

**Acknowledgements** The work in this paper is supported through the Sheafor/Lindsay Fund via the ERIT program at Rice’s Computer and Information Technology Institute CITI, NSF grant DMS-0240058, and NIBIB Grant No. 1T32EB006350-01A1

## Appendix

The following tables contain all the information pertaining to the ion channels and gating variable kinetics used in this paper. Table 2 is for the uniform channel model, which uses the Hodgkin–Huxley squid giant axon parameters. Table 3 is for the non-uniform channel model, whose non-uniformity comes from an A-type  $K^+$  current following Connor–Stevens kinetics and consistent with the graded channel distribution of (Hoffman et al. 1997).

**Table 2** Uniform channel model and kinetics, which corresponds to the Hodgkin–Huxley squid giant axon parameters at 6.3°C (Hodgkin and Huxley 1952)

Ionic Current	Gating Variables	$G(x)$ (mS/cm <sup>2</sup> )	$E$ (mV)	$\alpha$	$\beta$	$\tau$	$w_\infty$
Leak		0.3	−54.3				
$I_{Na}$	$m^3, h$	120	56				
$I_K$	$n^4$	36	−77				
	$m$			$\frac{0.1(v+40)}{1-\exp(-0.1(v+40))}$	$4\exp(-(v+65)/18)$	$\frac{1}{\alpha+\beta}$	$\frac{\alpha}{\alpha+\beta}$
	$h$			$0.07\exp(-(v+65)/20)$	$\frac{1}{1+\exp(0.1(-35-v))}$	$\frac{1}{\alpha+\beta}$	$\frac{\alpha}{\alpha+\beta}$
	$n$			$\frac{0.01(v+55)}{1-\exp(-0.1(v+55))}$	$\frac{1}{8}\exp(-(v+65)/80)$	$\frac{1}{\alpha+\beta}$	$\frac{\alpha}{\alpha+\beta}$



**Table 3** Non-uniform channel model and kinetics

Ionic current	Gating variables	$G(x)$ (mS/cm <sup>2</sup> )	$E$ (mV)	$\alpha$	$\beta$	$\tau$	$w_\infty$
Leak		0.3	-47				
$I_{Na}$	$m^3, h$	60	55				
$I_K$	$n^4$	20	-72				
$I_A$	$a^3, b$	$\frac{47.7}{2.3}(0.2 + 0.002x)$	-75	$\frac{0.38(v+29.7)}{1-\exp(-(v+29.7)/10)}$	$15.2\exp(-0.0556(v+54.7))$	$\frac{1}{\alpha+\beta}$	$\frac{\alpha}{\alpha+\beta}$
	$m$			$0.266\exp(-0.05(v+48))$	$\frac{3.8}{1-\exp(-(v+18)/10)}$	$\frac{1}{\alpha+\beta}$	$\frac{\alpha}{\alpha+\beta}$
	$h$			$\frac{0.02(v+45.7)}{1-\exp(-(v+45.7)/10)}$	$0.25\exp(-0.0125(v+55.7))$	$\frac{1}{\alpha+\beta}$	$\frac{\alpha}{\alpha+\beta}$
	$n$					$0.3632 + \frac{1.158}{1+\exp(0.0497(v+55.96))}$	$\left(\frac{0.0761\exp(0.0314(v+94.22))}{1+\exp(0.0346(v+1.17))}\right)^{1/3}$
	$a$					$1.24 + \frac{2.678}{1+\exp(0.0624(v+50))}$	$\left(\frac{1}{1+\exp(0.0688(v+53.3))}\right)^4$
	$b$						

$I_{Na}$  and  $I_K$  use Hodgkin–Huxley type kinetics, while  $I_A$  uses Connor–Stevens type kinetics and a spatial distribution of conductance based on Hoffman's work. For  $G(x)$ ,  $x$  is measured in  $\mu\text{m}$  from the soma. Note that the  $I_A$  channel kinetics do not have  $\alpha$  and  $\beta$  functions explicitly defined (Connor and Stevens 1971; Hoffman et al. 1997)

## References

- Antoulas, A. C., & Sorensen, D. C. (2001). Approximation of large-scale dynamical systems: An overview. *International Journal of Applied Math and Computer Science*, 11(5), 1093–1121.
- Ascoli, G. A. (2006). Mobilizing the base of neuroscience data: The case of neuronal morphologies. *Nature Reviews Neuroscience*, 7, 318–324.
- Brette, R., & Gerstner, W. (2005). Adaptive exponential integrate-and-fire model as an effective description of neuronal activity. *Journal of Neurophysiology*, 94, 3637–3642.
- Bush, P. C., & Sejnowski, T. J. (1993). Reduced compartmental models of neocortical pyramidal cells. *Journal of Neuroscience Methods*, 46, 159–166.
- Connor, J. A., & Stevens, C. F. (1971). Inward and delayed outward membrane currents in isolated neural somata under voltage clamp. *Journal of Physiology*, 213, 1–19.
- Fourcaud-Trocmé, N., Hansel, D., van Vreeswijk, C., & Brunel, N. (2003). How spike generation mechanisms determine the neuronal response to fluctuating inputs. *Journal of Neuroscience*, 23, 11628–11640.
- Gugercin, S., Antoulas, A., & Beattie, C. (2008).  $\mathcal{H}_2$  model reduction for large-scale linear dynamical systems. *SIAM Journal on Matrix Analysis and Applications*, 30, 609–638.
- Hasselmo, M. E., Giocomo, L. M., & Zilli, E. A. (2007). Grid cell firing may arise from interference of theta frequency membrane potential oscillations in single neurons. *Hippocampus*, 17, 1252–1271.
- Häusser, M. (2001). Synaptic function: Dendritic democracy. *Current Biology*, 11, R10–R12.
- Hines, M. (1984). Efficient computation of branched nerve equations. *International Journal of Bio-Medical Computing*, 15, 69–76.
- Hines, M. L., Morse, T., Migliore, M., Carnevale, N. T., & Shepherd, G. M. (2004). Modeldb: A database to support computational neuroscience. *Journal of Computational Neuroscience*, 17, 7–11.
- Hodgkin, A. L., & Huxley, A. F. (1952). A quantitative description of membrane current and its application to conduction and excitation in nerve. *Journal of Physiology*, 117, 500–544.
- Hoffman, D. A., Magee, J. C., Colbert, C. M., & Johnston, D. (1997).  $K^+$  channel regulation of signal propagation in dendrites of hippocampal pyramidal neurons. *Nature*, 387, 869–875.
- <http://NeuroMorpho.org>. The neuromorpho.org inventory. <http://NeuroMorpho.org>. Accessed March 11, 2008.
- <http://senselab.med.yale.edu/modeldb>. Senselab modeldb. <http://senselab.med.yale.edu/modeldb>. Accessed March 11, 2008.
- Hutcheon, B., & Yarom, Y. (2000). Resonance, oscillation and the intrinsic frequency preferences of neurons. *Trends Neuroscience*, 23, 216–222.
- Jolivet, R., Rauch, A., Lüscher, H.-R., & Gerstner, W. (2006). Predicting spike timing of neocortical pyramidal neurons by simple threshold models. *Journal of Computational Neuroscience*, 21, 35–49.
- Kailath, T. (1980). *Linear Systems*. Englewood Cliffs: Prentice Hall.
- Kistler, W. M., Gerstner, W., & van Hemmen, J. L. (1997). Reduction of the hodgkin–huxley equations to a single-variable threshold model. *Neural Computation*, 9, 1015–1045.
- Koch, C. (1999). *Biophysics of Computation*. Oxford: Oxford University Press.

- Magee, J. C., & Cook, E. P. (2000). Somatic epsp amplitude is independent of synapse location in hippocampal pyramidal neurons. *Nature Neuroscience*, 3, 895–903.
- Manwani, A., & Koch, C. (1999). Detecting and estimating signals in noisy cable structures, I: Neuronal noise sources. *Neural Computation*, 11, 1797–1829.
- Markram, H. (2006). The blue brain project. *Nature Rev. Neuroscience*, 7, 153–160.
- Martinez, J. O. (2008). Rice-baylor archive of neuronal morphology. Accessed May 1. <http://www.caam.rice.edu/~cox/neuromart>.
- Pinsky, P. F., & Rinzel, J. (1994). Intrinsic and network rhythmicogenesis in a reduced Traub model for CA3 neurons. *Journal of Computational Neuroscience*, 1, 39–60.
- Poznanski, R. (1991). A generalized tapering equivalent cable model for dendritic neurons. *Bulletin of Mathematical Biology*, 53(3), 457–467.
- Puil, E., Gimbarzevsky, B., & Miura, R. M. (1986). Quantification of membrane properties of trigeminal root ganglion neurons in guinea pigs. *Journal of Neurophysiology*, 55, 995–1016.
- Pyapali, G., Sik, A., Penttonen, M., Buzsaki, G., & Turner, D. (1998). Dendritic properties of hippocampal ca1 pyramidal neurons in the rat: Intracellular staining *in vivo* and *in vitro*. *Journal of Comparative Neurology*, 391, 335–352.
- Rall, W. (1959). Branching dendritic trees and motoneuron membrane resistivity. *Experimental Neurology*, 1, 491–527.
- Rudolph, M., & Destexhe, A. (2006). Analytical integrate-and-fire neuron models with conductance-based dynamics for event-driven simulation strategies. *Neural Computation*, 18, 2146–2210.
- Schierwagen, A. K. (1989). A non-uniform equivalent cable model of membrane voltage changes in a passive dendritic tree. *Journal of Theoretical Biology*, 141, 159–179.
- Thomson, A. M., & Lamy, C. (2007). Functional maps of neocortical local circuitry. *Frontiers in Neuroscience*, 1, 19–42.
- Timofeeva, Y., Cox, S. J., Coombes, S., & Josić, K. (2008). Democratization in a passive dendritic tree: An analytical investigation. *Journal of Computational Neuroscience*, 25, 228–244.
- Traub, R. D., & Miles, R. (1991). *Neuronal Networks of the Hippocampus*. Cambridge: Cambridge University Press.
- Trefethen, L. N., & Bau, D. (1997). *Numerical Linear Algebra*. Philadelphia: SIAM.
- Ulrich, D. (2002). Dendritic resonance in rat neocortical pyramidal cells. *Journal of Neurophysiology*, 87, 2753–2759.