ORIGINAL PAPER

The quantitative single-neuron modeling competition

**Renaud Jolivet · Felix Schürmann ·
Thomas K. Berger · Richard Naud ·
Wulfram Gerstner · Arnd Roth**

Abstract As large-scale, detailed network modeling
projects are flourishing in the field of computational neu-
roscience, it is more and more important to design single
neuron models that not only capture qualitative features of
real neurons but are quantitatively accurate in silico repre-
sentations of those. Recent years have seen substantial effort
being put in the development of algorithms for the systematic
evaluation and optimization of neuron models with respect
to electrophysiological data. It is however difficult to com-
pare these methods because of the lack of appropriate bench-
mark tests. Here, we describe one such effort of providing
the community with a standardized set of tests to quantify
the performances of single neuron models. Our effort takes
the form of a yearly challenge similar to the ones which have
been present in the machine learning community for some
time. This paper gives an account of the first two challenges
which took place in 2007 and 2008 and discusses future
directions. The results of the competition suggest that best
performance on data obtained from single or double electrode
current or conductance injection is achieved by models that
combine features of standard leaky integrate-and-fire models
with a second variable reflecting adaptation, refractoriness,
or a dynamic threshold.

Keywords Integrate-and-fire model · Quantitative
predictions · Benchmark testing · Scientific competition

R. Jolivet (✉)
Institute of Pharmacology and Toxicology,
University of Zurich, Zurich, Switzerland
e-mail: rjolivet@pharma.uzh.ch; renaud.jolivet@a3.epfl.ch

F. Schürmann · T. K. Berger · R. Naud · W. Gerstner
Brain Mind Institute, EPFL, Lausanne, Switzerland
e-mail: felix.schuermann@epfl.ch

T. K. Berger
e-mail: thomas.berger@epfl.ch

R. Naud
e-mail: richard.naud@epfl.ch

W. Gerstner
e-mail: wulfram.gerstner@epfl.ch

A. Roth
Wolfson Institute for Biomedical Research,
University College London, Gower Street,
London WC1E 6BT, UK
e-mail: arnd.roth@ucl.ac.uk

1 Introduction

Recent years have witnessed several important large scale
detailed network modeling projects being initiated in
computational neuroscience (Markram 2006; Izhikevich and
Edelman 2008). For these and other projects it is critical to
design single neuron models that can *quantitatively* capture
the activity of real neurons.

Single neuron models can be classified into two major
categories: detailed biophysical models and simple pheno-
menological models. Following the work of (Hodgkin and
Huxley 1952), a lot of effort has been spent to construct and
study biophysically realistic, detailed models of single neu-
ron electrical activity. These models can successfully repro-
duce a large variety of neuronal behaviors as observed in
experiments by an appropriate combination of different ion
currents (Bower and Beeman 1995). The development of
algorithms for automatic fitting of such models to data in
order to obtain models with quantitatively predictive power
is an active area of research (Prinz et al. 2003, 2004; Achard
and De Schutter 2006; Huys et al. 2006; Druckmann et al.
2007; Keren et al. 2005; Bush et al. 2005; Gerken et al. 2006;
Vanier and Bower 1999; Weaver and Wearne 2006).

In contrast to detailed Hodgkin–Huxley-type models, very simple models have only a small number of parameters which can be automatically and easily extracted from electrophysiological recordings. In the 1970s and 1980s, researchers started to develop methods for the extraction of simple neuron models from neural data. Brillinger and Segundo, in particular, used maximum-likelihood and optimal filtering techniques to evaluate the linear response curve and firing probability of neuronal membranes, laying down the foundations for more contemporary approaches (Brillinger and Segundo 1979; Brillinger 1988a,b). Recently, several groups have started to build models with true quantitative predictive power. Rauch, La Camera, and colleagues have demonstrated that the output frequency of cortical pyramidal neurons and interneurons recorded in vitro can be fitted by Integrate-and-Fire neurons (La Camera et al. 2004; Rauch et al. 2003). Keat and colleagues have shown that the precise spike timing of neurons recorded extracellularly in the visual pathway can be predicted almost exactly with a very simple model neuron (Keat et al. 2001). Similar results have been obtained for retinal ganglion cells by Pillow et al. (2005). Carandini and colleagues have shown that spike trains of a neuron in the mammalian brain can be predicted on the basis of the spike train of its main afferent (Carandini et al. 2007). Several techniques have been successfully applied to predict the membrane voltage and spike timing of cortical pyramidal neurons recorded intracellularly in vitro (Clopath et al. 2007; Jolivet et al. 2006b; Paninski et al. 2005; Badel et al. 2008) and in vivo (Lansky et al. 2006). Finally, several methods have been proposed that still await testing on experimental data (Jolivet and Gerstner 2004; Kobayashi and Shinomoto 2007; Mullowney and Iyengar 2008).

Simple models can qualitatively reproduce a broad range of observed neuronal behaviors (Izhikevich 2004). Their simplicity permits one to mathematically analyze questions of neural coding (Hopfield 1995; Gerstner et al. 1996; Kempter et al. 1998; Gerstner and Kistler 2002; Arcas and Fairhall 2003; Brunel et al. 2003; Keat et al. 2001; Pillow et al. 2005; Leibold et al. 2008) and implementation in neural prostheses where such models can be simulated at low cost or built in silico (Song et al. 2007; Marmarelis and Berger 2005).

Despite this intense activity, the community still lacks benchmark tests that can be used as a reference to compare cost and performance of different methods. Here, we describe one such effort of providing the community with a standardized set of tests to quantify the performance of single neuron models. Our effort takes the form of a yearly challenge similar to those in the machine learning community. This paper gives an account of the first two challenges which took place in 2007 and 2008 and discusses future directions.

2 The quantitative single-neuron modeling competition

The quantitative single-neuron modeling competition is divided into four individual challenges, each one addressing different aspects of neuronal modeling. Challenges A and B are focused on single-compartment models while challenges C and D are focused on multicompartmental models. The goal, dataset and associated evaluation criteria are briefly described below for each of the four challenges. The data of challenge C and D were identical in 2007 and 2008. Challenge A and B were modified and replaced with data exclusively recorded for the challenge in 2008. Datasets and evaluation criteria are subject to minor changes in subsequent competitions in order to improve the benchmarking. Please visit the competition's website for further details about the present competition.[1]

2.1 Challenge A: predicting the timing of somatic action potentials (2007)

Challenge A proposes to develop a quantitative neuron model with the goal of predicting the timing of output spikes with ± 2 ms precision. We chose this precision level because it is of the same order than synaptic rise times that can be measured in the soma of cortical pyramidal neurons. Data were obtained from a cortical pyramidal neuron in slices of rat barrel cortex [data courtesy of Alexander Rauch and colleagues, see (Rauch et al. 2003; Jolivet et al. 2006b) for details of the experimental procedures]. Stimulus currents were injected and voltage responses were measured via whole-cell patch-clamp recording at the soma. Randomly fluctuating currents with different mean and variance were used as stimulus waveforms, and each stimulation was repeated four times in order to evaluate the intrinsic reliability of the cell in terms of the timing of output spikes (Fig. 1). The fluctuating currents were generated offline once. The same random sequence was then injected at each repetition.

Data are divided into two sets. One set, the so-called training set, is made available to the public and contains both the injected current and the voltage recorded in all four trial repetitions. In the second set, the so-called test set, only the driving current is made available to the public. Contestants simulate their quantitative model with these currents as input and submit their predicted output spike times that are then independently evaluated by the organizers.

Evaluation criterion. The aim is to predict spike times with a precision of $\Delta = \pm 2$ ms. To evaluate the quality, we calculate the number of coincidences N_{coinc} between the spikes in the data spike train (target) and the spike train of the tested model.
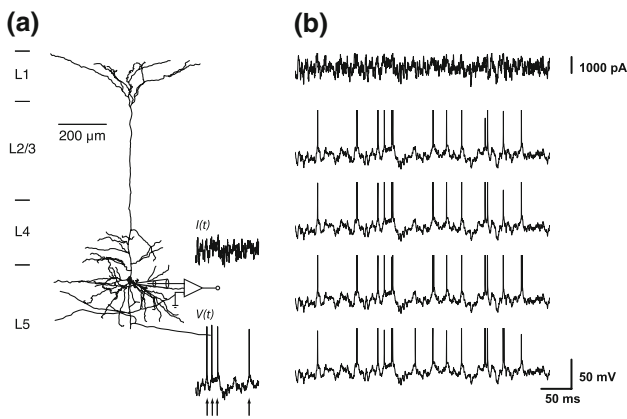

Fig. 1 Challenge A 2007. **a** A cortical pyramidal neuron was stimulated by injection of randomly fluctuating currents of various amplitudes. Current was injected and voltage responses were recorded at the soma. The goal of the challenge is to predict the timing of somatic action potentials with ±2 ms precision, given the injected current as a function of time. **b** A sample of injected current (*top*) and waveforms in order to evaluate the intrinsic reliability of the cell in terms of the timing of output spikes Jolivet et al. (2006b)

We subtract the expected number of coincidences $\langle N_{\text{coinc}} \rangle$ that a spike train generated by a Poisson process with rate $\nu = N_{\text{data}}/T$ (where T is the duration of the experiment) would give and we divide by the number of spikes in the two spike trains. Here,

$$\Gamma = \alpha \frac{N_{\text{coinc}} - \langle N_{\text{coinc}} \rangle}{\frac{1}{2}(N_{\text{data}} + N_{\text{model}})} \qquad (1)$$

where N_{data} and N_{model} denote the number of spikes in the data and model spike trains and $\alpha = 1/[1 - 2\nu\Delta]$ is a factor that normalizes the coincidence factor Γ to a maximum of 1. $\Gamma = 0$ implies that the prediction is not better than chance level. $\Gamma = 1$ implies that the prediction by the model is optimal.

The overall aim is to maximize Γ averaged across all test sets and normalize by the intrinsic reliability of the cell for each particular dataset. For each input, four responses of the neuron are available (the same input was repeated four times). We calculate $\Gamma(k)$ for all n combinations of model response and neuron response $1 \le k \le n$ as well as the intrinsic reliability Γ_{int} and return

$$\Gamma_A = \frac{1}{n} \sum_k \frac{\Gamma(k)}{\Gamma_{\text{int}}} \qquad (2)$$

Here Γ_{int} is the factor Γ as in Eq. (1) but calculated between two experimental spike trains obtained from two repetitions of the same stimulus, rather than between one model spike train and one experimental spike train. Note that several other measures exist that go beyond firing rates and consider precise firing times. Some measures are based on binning of the spike trains (Geisler et al. 1991; MacLeod et al. 1998) or on

cost functions (Aronov and Victor 2004; van Rossum 2001; Victor and Purpura 1996, 1997). The present measure was originally proposed by Kistler et al. (1997) and used in several previous studies (Clopath et al. 2007; Jolivet et al. 2004, 2006b; Jolivet and Gerstner 2004). See (Jolivet et al. 2008) for further details.

2.2 Challenge A: predicting the timing of somatic action potentials (2008)

In 2008, this dataset was replaced by data from cortical neurons under random conductance injection, which moved the input paradigm closer to the *in-vivo* situation. Conductance was injected via dynamic current clamping with two somatic electrodes, one electrode delivering the voltage-dependent current and the other one serving as a voltage follower. Excitatory (g_{exc}) and inhibitory (g_{inh}) noisy conductances are simultaneously injected with correlation times of τ_{exc} and τ_{inh}, respectively. The injected net current is therefore

$$I(t) = g_{\text{exc}}(t)[V(t) - E_{\text{exc}}] + g_{\text{inh}}(t)[V(t) - E_{\text{inh}}] \qquad (3)$$

with reversal potentials of $E_{\text{exc}} = -10$ mV and $E_{\text{inh}} = -70$ mV and where $g_k(t)$ ($k = \text{exc}$ or $k = \text{inh}$) is obtained from

$$\tau_k \frac{dg_k}{dt} = -[g_k - \mu_k] + a_k \xi(t) \qquad (4)$$

where τ_k is the correlation time ($\tau_{\text{exc}} = 2$ ms and $\tau_{\text{inh}} = 10$ ms, μ_k is the mean conductance, and $\xi(t)$ is white noise. The mean values of excitatory and inhibitory conductances are always $\mu(g_{\text{ex}}) = 12$ nS and $\mu(g_{\text{inh}}) = 5$ nS, but the standard deviations of the excitatory and inhibitory signals are different for each stimulus (Fig. 2). The training set consists of six different stimuli, and three different stimuli serve as test
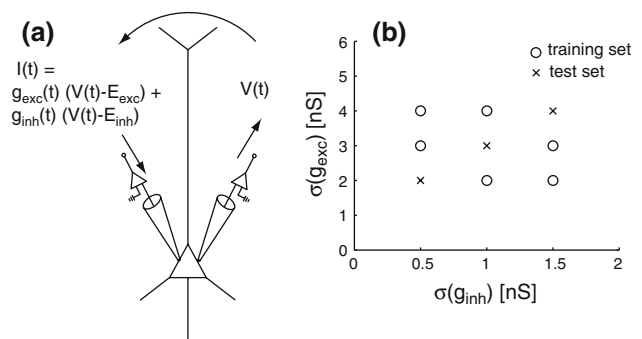

Fig. 2 Challenge A 2008. **a** A layer 5 cortical pyramidal neuron was injected with randomly fluctuating excitatory and inhibitory conductances. Current was injected via one electrode as a function of the instantaneously measured voltage at another electrode. Once again, the goal was to predict the timing of outgoing spikes with a ±2 ms precision knowing the injected conductances. **b** Combinations of standard deviations of excitatory and inhibitory conductances that have been injected

Fig. 3 Challenge B 2007.
a Two cortical neurons, a
pyramidal cell and a bitufted
cell (not shown), were injected
with various types of ramp and
step currents. Voltage was
recorded at the soma. The goal
of the challenge is to predict the
subthreshold voltage trace, the
timing of the first spike, as well
as the firing frequency at the
beginning and at the end of a
two-second step current (see
text for further details). **b** Four
samples of injected current (*top*)
and of recorded voltage (*bottom*)
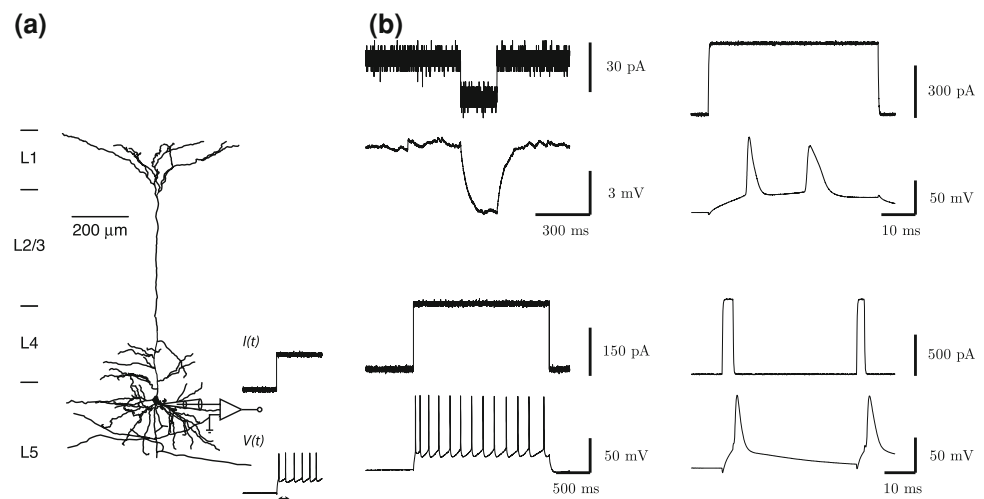
sets. Each stimulus is repeated 7–8 times. Further information
on the experimental methods can be found on the competi-
tion's webpage (see Footnote 1).

2.3 Challenge B: predicting the neuronal response upon stimulation by ramp and step currents (2007)

Challenge B has the goal to predict the response of neurons
upon stimulation by ramp and step currents. Data were col-
lected in a cortical pyramidal neuron and a cortical bitufted
cell from a young rat's (postnatal day 13–15) primary soma-
tosensory, non-barrel cortex [data courtesy of the Blue Brain
Project[2] and see Wang et al. (2002) for details on the expe-
rimental procedures]. Neurons were injected with various
types of artificial stimuli and voltage was recorded at the
soma in current-clamp mode (Fig. 3).

Several input samples are given as training data toge-
ther with the electrophysiological measurements. Other input
samples are given for test/prediction (without the corres-
ponding measurements). More specifically, the training set
consists of

– Hyperpolarizing and depolarizing subthreshold current
  steps (Fig. 3b top left).
– Superthreshold step currents of about 0.1 s duration
  (Fig. 3b top right). Cells respond with a few spikes.
– Superthreshold step currents of about 2 s duration (Fig. 3b
  bottom left). One cell responds with spike trains that show
  adaptation; another cell exhibits initial bursting without
  further adaptation.
– Pairs of strongly depolarizing pulses that allow one to
  study spike-afterpolarizing potentials, spike triggered
  adaptation and refractoriness (Fig. 3b bottom right).

The test set consists of

– Two sequences of different subthreshold steps, one
  depolarizing, the other hyperpolarizing.
– Superthreshold step currents of about 0.2 s duration, simi-
  lar to the one in the training set, but starting from a depo-
  larized background.
– A slow ramp current leading to neuronal firing.

Evaluation criterion. The aim is to predict the subthreshold
voltage trace, the timing of the first spike, as well as the firing
frequency at the beginning and at the end of a two-second
step current. Model quality is measured first separately for
each criterion, and then the overall rating is determined by
appropriate mixing of the individual criteria. Each criterion
is evaluated on a scale between 0 and 1 where 1 is the optimal
solution. In addition to the coincidence factor Γ of challenge
A, we use a generic function

$$g(x) = \frac{1}{1 + x^2} \tag{5}$$

which is bounded between 0 and 1. Equation (5) is used as
follows on the different test sets mentioned above

– Two sequences of different subthreshold steps, one depo-
  larizing, the other hyperpolarizing. The aim is to predict
  the subthreshold voltage with a precision of 2 mV. We
  define the difference between the voltage of the experi-
  mental trace and that of the model neuron in data set k
  as $x_k(t) = (u_{\text{data},k}(t) - u_{\text{model},k}(t))/2$ mV and use the
  criterion

$$F_1 = \frac{1}{2} \sum_{k=1}^{2} \frac{1}{T} \int_0^T g(x_k(t)) dt \tag{6}$$

where k runs over the two sequences of subthreshold steps and T is the duration of the stimulation.

- Several superthreshold step currents of about 0.2 s duration starting from depolarized background. The aim is to predict the timing of the first spike with a precision of 10 ms. We define the difference between the timing of the first spike of the experimental neuron and that of the model neuron as $x^{\text{first}} = (t_{\text{data}}^{\text{first}} - t_{\text{model}}^{\text{first}})/10$ ms and use the criterion

$$F_2 = \frac{1}{n} \sum_{k=1}^{n} g(x_k^{\text{first}}) \tag{7}$$

The lower index refers to the different data sets (step currents of different amplitudes).

- The second aim with these superthreshold step currents is to predict the firing frequency of the neuron at the beginning and at the end of the stimulation with a precision of 5 Hz. To do so, we determine the duration of the first, second, and last interspike interval and transform to units of frequency. For example, the frequency during the first interval is the inverse of the first interspike interval $f^{\text{first}} = 1/[t^{\text{second}} - t^{\text{first}}]$. The normalized difference between the frequency of the experimental neuron and that of the model neuron $x^{f1} = (f_{\text{data}}^{\text{first}} - f_{\text{model}}^{\text{first}})/5$ Hz leads to the evaluation measure $g(x^{f1})$. Similarly, the inverse of the duration of the second and last interspike intervals leads to measures $g(x^{f2})$ and $g(x^{f3})$. The final evaluation criterion is

$$F_{i+2} = \frac{1}{n} \sum_{k=1}^{n} g(x_k^{fi}) \tag{8}$$

with $i = 1, 2, 3$. The index k runs over all experimental traces of step currents in the test set.

- A slow ramp current, leading to neuronal firing. The aim is to predict the timing of the first spike with a precision of 10 ms. We define the difference between the timing of the first spike of the experimental neuron and that of the model neuron as $x^{\text{first}} = (t_{\text{data}}^{\text{first}} - t_{\text{model}}^{\text{first}})/10$ ms and use the criterion

$$F_6 = g(x^{\text{first}}) \tag{9}$$

The overall aim is to maximize criteria $F_1, \ldots, F_6$ for each of the two different neurons. Since we have two neurons (regular spiking with adaptation and delayed spike initiation), we have a total of 12 criteria. We use the labels $F_1, \ldots, F_6$ for the first neuron and $F_7, \ldots, F_{12}$ for the second neuron. In the literature of multi-criteria optimization several possibilities of combining the different "targets" are discussed. A particularly severe one is to look only at the worst performance.

An alternative would be to look at the mean performance $\sum_k F_k/12$. We use a compromise between both approaches and average over all 12 criteria but give more importance to the worst performing criteria. With this approach, we still consider all 12 criteria but avoid the situation that a model ranks high if it does very well on the majority of criteria (say 9 out of 12) but performs really poorly (with $F_k \approx 0$) on the remaining ones. Loosely speaking, we want to give preference to models that perform reasonably well across a broad range of stimuli. Specifically, we use the following two-step procedure

1. Rank the indices according to performance. The criteria that performs worst gets an upper index (1): $C^{(1)} = \min_k\{F_k\}$, the second worst an upper index (2): $C^{(2)} = \min_k\{F_k|\text{worst removed}\}$, etc . . .
2. Evaluate the weighted mean

$$\Gamma_B = 0.5\, C^{(1)} + 0.25\, C^{(2)} + \cdots = \sum_{k=1}^{12} \frac{1}{2^k} C^{(k)} \tag{10}$$

The overall aim is to maximize Γ_B.

2.4 Challenge B: predicting the neuronal response upon stimulation by ramp and step currents (2008)

The challenge B in 2008 uses fewer (three) training stimuli and only one test stimulus. The number of repetitions was increased in order to have a better measure of the cell's intrinsic reliability. The training stimuli consisted of depolarizing steps, hyperpolarizing steps, and a symmetric ramp stimulation (Fig. 4). The test stimulus consisted of a concatenation of de- and hyperpolarizing steps.

Evaluation criterion. The evaluation is performed on 18 different objectives which have to be satisfied simultaneously by the same model. The features consist of the following: total number of spikes, inter-spike intervals (ISI), spike latency (SL), subthreshold voltage (SV), and inter-spike minimum (ISM) of the voltage trace. The test stimulus consists of several epochs [see Fig. 4]. Some of the features do not apply to all epochs, and it would be overly complicated to consider all the possible features of the neuronal response. Therefore, we focus on the following restricted set of 18 features, namely

- f_1: total number of spikes
- f_2: first SL in epoch 2
- f_3: first ISI in epoch 2
- f_4: second ISI in epoch 2
- f_5: last ISI in epoch 2
- f_6: mean ISM in epoch 2
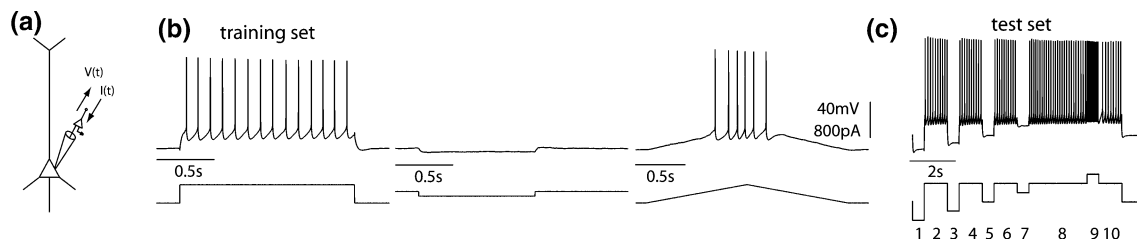- f_7: SV in the initial segment of epoch 2

Fig. 4 Challenge B 2008. **a** Cortical pyramidal neurons were stimulated by injection of current traces at the soma. Voltage was recorded with the same electrode. **b** 3 different training stimuli were used: a suprathreshold depolarizing step, a hyperpolarizing step, and a symmetric ramp. **c** The test stimulus consisted of a concatenation of step currents of various amplitudes. Ten different epochs of this stimulation are labeled. The total number of spikes and features from the epochs 2, 3, 8, 9, and 10 are evaluated

- f_8: SV in epoch 3
- f_9: first SL in epoch 8
- f_{10}: first ISI in epoch 8
- f_{11}: second ISI in epoch 8
- f_{12}: last ISI in epoch 8
- f_{13}: mean ISM in epoch 8
- f_{14}: SV in the initial segment of epoch 8
- f_{15}: first ISI in epoch 9
- f_{16}: last ISI in epoch 9
- f_{17}: first ISI in epoch 10
- f_{18}: last ISI in epoch 10

For each feature, we compute the chi-square deviation in order to get the set of 18 performance measures. For example, the first performance measure is

$$B_1 = \sqrt{\frac{(N_m - \langle N_n \rangle)^2}{V[N_n]}} \tag{11}$$

where N_m is the number of spikes predicted by the model, N_n the number of spikes observed in the real neuron. The angular brackets stand for the average over all available repetitions, and V[.] stands for the variance across all available repetitions of the experimental trace. A slightly different procedure is used for the features concerned with the subthreshold voltage. We estimate the mean square difference between the voltage measured and the voltage modeled, at each time step from the onset of the current step to a time T. This is weighted by the average mean squared error across all combinations of the six repetitions. The extent of the initial segment considered, T, is chosen to be the first spike latency minus three times its standard deviation.

2.5 Challenge C: predicting the timing of somatic action potentials in cortical neurons in response to random current injection at the soma and a dendritic location (2007 and 2008)

Challenge C asks for a quantitative neuron model that is able to predict the response of neurons to random current injection

at the soma and/or its dendrites (Fig. 5). The dataset consists of simultaneous whole-cell voltage recordings from the soma and apical dendrite of neocortical layer 5 pyramidal neurons under random current injection via the somatic and/or dendritic recording pipette [data courtesy of Matthew Larkum and colleagues (Larkum et al. 2004)]. Stimulation is done with currents of different means and fluctuation amplitudes.

For each combination of somatic and dendritic inputs, only a single trial is available in the training set. As for challenges A and B, several stimulation sets are set apart as test sets for the prediction.

Evaluation criterion The aim is to predict the timing of somatic action potentials with a precision of ± 2 ms. To evaluate the quality, we use the same criterion as in challenge A, i.e., $\Gamma_C = \Gamma$ [see Equation (1)] The overall aim is to maximize Γ averaged across all test sets.

2.6 Challenge D: predicting the timing of somatic action potentials and the evolution of the dendritic membrane potential in a layer 5 cortical pyramidal neuron in response to current injection at the soma and up to 2 dendritic locations (2007 and 2008)

Challenge D combines features of challenges B and C and focuses on voltage traces obtained by triple whole-cell recordings from the soma and two dendritic locations in a neocortical layer 5 pyramidal neuron [data courtesy of Matthew Larkum and colleagues (Larkum et al. 2004)]. Two different types of stimulation protocols are used. In the first, current pulses are injected with different delays at the soma and at the distal dendritic pipette to capture the nonlinear interactions between backpropagating action potentials and dendritic calcium spikes during BAC firing (backpropagating action potential activated calcium spike firing; Fig. 6) (Larkum et al. 1999; Schaefer et al. 2003). In the second, random current waveforms are injected at the soma and/or at dendritic locations, similar to those used in challenge C.

Fig. 5 Challenge C. **a** Cortical pyramidal neurons were stimulated by injection of random current waveforms at the soma and/ or dendrites. Voltage was recorded at the somatic and/or dendritic site. The goal is to predict the timing of somatic action potentials with a precision of ± 2 ms, as in challenge A. **b** Samples of injected current (*top*) and of recorded voltage (*bottom*) at the somatic site (subscript 's', black traces) and at the dendritic site (subscript 'd', dotted red traces)
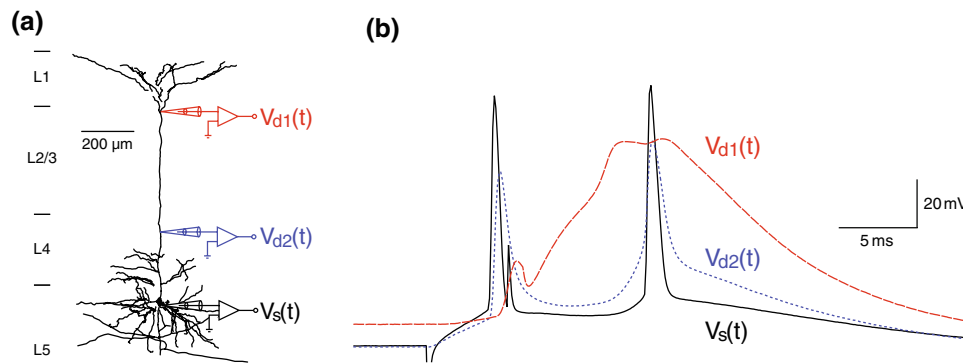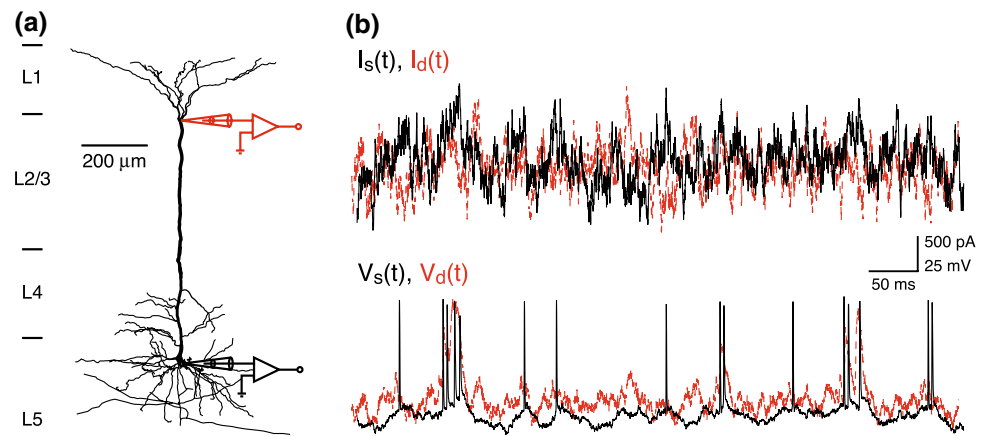

Fig. 6 Challenge D. **a** Recording configuration used to evoke and observe backpropagating action potential activated calcium spike firing [BAC firing; (Larkum et al. 1999)]. At the soma, a square pulse of current (duration, 5 ms; amplitude, 2.5 nA) is injected in each trial via the patch pipette (*solid black*). At the proximal dendritic pipette (*dotted blue*) zero current is injected. At the distal dendritic pipette (*dashed red*) an EPSC-like current waveform is injected (time course,

$f(t) = (1 - \exp(-t/\tau_1)) * \exp(-t/\tau_2)$ with $\tau_1 = 2$ ms and $\tau_2 = 8$ ms; peak amplitude of the resulting waveform, 2.6 nA) whose onset is varied with respect to the time of the somatic current pulse. The membrane potential is recorded simultaneously at all three pipettes, as shown in **b** for a trial in which the distal dendritic EPSC waveform began 1 ms after the end of the somatic current pulse

Evaluation criterion The first aim is to predict the trajectory of the membrane potential at the location of the proximal dendritic pipette during the BAC firing protocol with a precision of 2 mV. The voltage at this pipette is recorded faithfully because it is not used for injecting current during this protocol. We define the difference between the voltage of the experimental trace and that of the model neuron in data set k as $x_k(t) = (u_{\text{data},k}(t) - u_{\text{model},k}(t))/2$ mV and use the criterion

$$F_1 = \frac{1}{n} \sum_{k=1}^{n} \frac{1}{T} \int_0^T g(x_k(t)) dt \qquad (12)$$

where k runs over all n sequences of subthreshold steps and T is the duration of the stimulation. Similar to challenge B, the function g is bounded between 0 and 1 and given by $g(x) = 1/(1 + x^2)$ in challenge D 2007 and $g(x) = 1/(1 + |x|)$ in challenge D 2008, respectively. The second aim is to predict

the timing of somatic action potentials in response to both types of stimulation protocols with a precision of ± 2 ms. To evaluate the quality of the prediction, we use the same coincidence measure that is used for challenges A and C. F_2 is the value of Γ averaged over all samples in the test set. Finally, both the subthreshold criterion F_1 and the spike precision criterion F_2 will be given equal weight

$$\Gamma_D = 0.5 F_1 + 0.5 F_2 \qquad (13)$$

The overall aim is to maximize Γ_D.

3 Discussion

In this paper, we briefly introduced the *Quantitative Single-Neuron Modeling Competition* as it took place in 2007 and 2008. Given the feedback we had from the first two editions (see Tables 1 and 2), we plan to make it an annual scientific

Table 1 Challenge A 2007

Authors	Performance	Model
Benchmark	0.82	AdEx
Kobayashi and Shinomoto	0.72	ARX
Benchmark	0.69	aSRM
Shinomoto and Kobayashi	0.66	CC
Anonymous	0.55	
Anonymous	0.32	

Max. performance is 1. AdEx: adaptative exponential Integrate-and-Fire model; ARX: Autoregressive exogenous model; aSRM; adaptative Spike Response Model; CC: carbon-copy model [see (Jolivet et al. 2008) for further details]

Table 2 Challenge A 2008

Authors	Performance	Model
Shinomoto and Kobayashi	91.4 ± 7.1	ARX
Badel	89.9 ± 6.0	AdEx-1
Mensi	84.4 ± 4.4	aSRM
Mueller	84.2 ± 4.2	AdEx-2
Anonymous	82.1 ± 4.4	
Anonymous	73.9 ± 7.3	
Benchmark	48.2 ± 2.0	LIF
Guerel-Ponulak	46.9 ± 4.5	
Anonymous	37.4 ± 1.4	
Anonymous	33.3 ± 2.7	
Anonymous	18.4 ± 0.8	
Anonymous	13.6 ± 1.4	

Max. performance is 100. ARX: Autoregressive exogenous model with moving threshold; AdEx: exponential Integrate-and-Fire with dynamic threshold; aSRM: Spike Response Model with moving threshold; LIF: leaky Integrate-and-Fire model

event during the next couple of years. Of course, the datasets and evaluation criteria are subject to improvement over the next years. For example, in the 2008 edition of the challenge we switched in Challenge A from a random current to a random conductance injection which reflects more closely the situation in vivo. Moreover, for challenge B in 2008 we used a different set of recordings with multiple repetitions for each step-stimulus so as to allow to scale the performance of a neuron model on a given criterion (such as spike timing, frequency, or subthreshold voltage) with the intrinsic reliability of the experimental neuron for that feature.

3.1 What is a good evaluation criterion?

Designing appropriate evaluation criteria is one of the most important issues of running a competition. While it is relatively easy to develop an elegant evaluation criterion to compare the timing of spikes (see Γ_A for instance), it is very difficult to develop an evaluation criterion for predicting the membrane voltage or other features of neuronal dynamics that might be of importance. A naive approach based on the squared difference of predicted and measured voltage is not useful because of the singular character of spike events. An action potential misplaced by one millisecond would cause a large error signal. However, the form of the downswing of the spike and the spike-afterpotential have been used by experimentalists to distinguish different neuron types, and here small differences in the voltage matter. Hence we wanted to include some features characterizing these aspects. Finally, traditional electrophysiology has characterized neurons by spike frequency and spike frequency adaptation. The combination of all these aspects led to the set of rather complex criteria used in challenge B.

An interpretation of the raw values of the different test criteria might prove difficult: Is 0.98 really much better than 0.90? Assessing the quality of a model on a specific criterion is not possible without comparing it with the intrinsic reliability of the experimental neuron on this same criterion Jolivet et al. (2006b) [or a class of neurons as discussed in Druckmann et al. (2007)]. In 2008, we have therefore used in challenge B measures that are scaled by the intrinsic reliability, that is, our criteria tell us by how many experimental standard deviations a model differs from the target. This approach allows us to compare numbers across different criteria. It should be noted that a scaling by intrinsic reliability is only possible if a sufficient number of repetitions of each experimental paradigm is available—and this requires a careful design of the experimental protocol. A lot of work is still needed to develop and characterize appropriate evaluation functions that quantify model predictions beyond a simple average firing rate or coincident spikes.

3.2 What is a good neuron model?

Challenges A and B probe the neuron at a single location and therefore relate naturally to single-compartment models (point neuron models), whereas challenges C and D require at least two compartments. Although many researchers expressed interest in the challenges, only a small number of submissions was received. Submissions were especially scarce for challenges C and D. Challenge A appeals directly to the large community of researchers interested in single-compartment models – and because of the simple evaluation criteria (spike timing) it seems to be amenable to standard optimization techniques. Surprisingly, challenge B which takes up electrophysiological paradigms used by numerous researchers did not incite a large number of submissions, neither with detailed models of the Hodgkin-Huxley type nor with simple integrate-and-fire type models. Challenges C-D represent a more difficult problem whose solution requires models consisting of more than one compartment. However,

we predict that significant progress will be made in this area at the interface between simple phenomenological models and detailed biophysical models in the near future. In summary, we believe that a lot can be learned about neuronal computation by taking up this kind of challenge, and we would like to encourage the computational neuroscience community to participate in subsequent competitions in truly sportive spirit.

What follows from the results of challenge A displayed in Tables 1 and 2 is that standard leaky integrate-and-fire models or other off-the-shelf methods are not sufficient to account for the variety of firing patterns and firing rates generated by a single neuron. The conclusion is that one has to include some dynamics in the threshold so as to achieve two things: first, to account in some rough fashion for neuronal refractoriness, and, second, to gain some flexibility in matching the mean firing rates across different stimulation paradigms. We had already shown that predicting subthreshold membrane voltage is relatively easy (Jolivet et al. 2006a). Predicting the exact timing of spikes is where the difficulty resides.

In order to reach a substantial predictive capability over a broad range of firing rates, some of us and others had to include dynamics in the threshold mimicking adaptation and/or refractoriness (Rauch et al. 2003; Jolivet et al. 2006b). Indeed, all the models that included one such feature performed well. ARX and aSRM (see Tables 1 and 2) have a cumulative dynamic threshold. Apart from the dynamic threshold, both models are purely linear. The ARX performed slightly better than the aSRM since the ARX submission used a different set of parameters for each test set whereas the aSRM used the same set of parameters across all stimuli. The AdEx model [adaptive exponential integrate-and-fire model (Brette and Gerstner 2005)] combines a nonlinear voltage equation containing linear and exponential terms (Fourcaud-Trocmé et al. 2003) with a second equation that captures both spike-triggered adaptation and subthreshold adaptation. Two independent submissions used the AdEx models with some slight modifications. The submission of Badel (AdEx-1) combined the AdEx dynamics with a moving threshold; the submission of Mueller (AdEx-2) modified one term in the AdEx equations from current based to conductance based. All these models with adaptation and/or dynamic threshold performed very well. The difference between the ARX and the AdEx-1 was not significant so that both can be considered as winners of the 2008 competition. On the 2007 data set, the AdEx model ran out of competition, but performed slightly better than the ARX model. In contrast to the models mentioned so far, all those models that did not include features such as adaptation of dynamic threshold had a significantly lower performance (most of the anonymous results). In particular, an off-the-shelf leaky integrate-and-fire model achieved a performance of 48.2 ± 2 on the 2008 data set, which is several standard deviations below the performance of the winning adaptive models which score in the range of 90 ± 7. The exact optimization strategy does not seem to play a critical role. A detailed description of the first four models and associated optimization strategies presented in Table 1 can be found in (Jolivet et al. 2008). Winning strategies were very similar in both challenge years.

Note that the present initiative differs from the recent Neural Prediction Challenge[3] in the sense that the goal is to design a model that predicts features of neuronal activity in response to fluctuating current or conductance inputs while the goal in the Neural Prediction Challenge is to design a model that predicts the responses of neurons to "natural" sensory stimuli in vivo.

Further information about the subsequent and future challenges can be found on the competition's website (see Footnote 1).

References

Achard P, De Schutter E (2006) Complex parameter landscape for a complex neuron model. PLoS Comp Biol 2(7):e94

Arcas B, Fairhall A (2003) What causes a neuron to spike. Neural Comp 15:1789–1807

Aronov D, Victor JD (2004) Non-euclidean properties of spike train metric spaces. Phys Rev E 69:061,905

Badel L, Lefort S, Brette R, Petersen CCH, Gerstner W, Richardson MJE (2008) Dynamic IV curves are reliable predictors of naturalistic pyramidal-neuron voltage traces. J Neurophysiol 99(2): 656–666

Bower JM, Beeman D (1995) The book of Genesis. Springer, New York

Brette R, Gerstner W (2005) Adaptive Exponential Integrate-and-Fire Model as an Effective Description of Neuronal Activity

Brillinger DR (1988a) Maximum likelihood analysis of spike trains of interacting nerve cells. Biol Cybern 59:189–200

Brillinger DR (1988b) The maximum likelihood approach to the identification of neuronal firing systems. Ann Biomed Eng 16:3–16

Brillinger DR, Segundo JP (1979) Empirical examination of the threshold model of neuronal firing. Biol Cybern 35:213–220

Brunel N, Hakim V, Richardson MJE (2003) Firing-rate resonance in a generalized integrate-and-fire neuron with subthreshold resonance. Phys Rev E 67:051,916

Bush K, Knight J, Anderson C (2005) Optimizing conductance parameters of cortical neural models via electrotonic partitions. Neural Net 18(5–6):488–496

Carandini M, Horton JC, Sincich LC (2007) Thalamic filtering of retinal spike trains by postsynaptic summation. J Vision 7(14):20

Clopath C, Jolivet R, Rauch A, Lüscher HR, Gerstner W (2007) Predicting neuronal activity with simple models of the threshold

---

[3] http://neuralprediction.berkeley.edu/.

type: Adaptive Exponential Integrate-and-Fire model with two compartments. Neurocomputing 70(10–12):1668–1673

Druckmann S, Banitt Y, Gidon A, Schürmann F, Markram H, Segev I (2007) A novel multiple objective optimization framework for constraining conductance-based neuron models by experimental data. Front Neurosci 1:7–18

Fourcaud-Trocmé N, Hansel D, van Vreeswijk C, Brunel N (2003) How spike generation mechanisms determine the neuronal response to fluctuating inputs. J Neurosci 23:11,628–11,640

Geisler WS, Albrecht DG, Salvi RJ, Saunders SS (1991) Discrimination performance of single neurons: rate and temporal information. J Neurophysiol 66:334–362

Gerken WC, Purvis LK, Butera RJ (2006) Genetic algorithm for optimization and specification of a neuron model. Neurocomputing 69(10–12):1039–1042

Gerstner W, Kistler WM (2002) Spiking neuron models. Cambridge University Press, Cambridge

Gerstner W, Kempter R, van Hemmen JL, Wagner H (1996) A neuronal learning rule for sub-millisecond temporal coding. Nature 383:76–78

Hodgkin AL, Huxley AF (1952) A quantitative description of membrane current and its application to conduction and excitation in nerve. J Physiol (London) 117:500–544

Hopfield JJ (1995) Pattern recognition computation using action potential timing for stimulus representation. Nature 376:33–36

Huys QJM, Ahrens MB, Paninski L (2006) Efficient estimation of detailed single-neuron models. J Neurophysiol 96:872–890

Izhikevich EM (2004) Which model to use for cortical spiking neurons. IEEE Trans Neural Net 15:1063–1070

Izhikevich EM, Edelman GM (2008) Large-scale model of mammalian thalamocortical systems. PNAS 105(9):3593–3598

Jolivet R, Gerstner W (2004) Predicting spike times of a detailed conductance-based neuron model driven by stochastic spike arrival. J Physiol-Paris 98(4–6):442–451

Jolivet R, Lewis TJ, Gerstner W (2004) Generalized integrate-and-fire models of neuronal activity approximate spike trains of a detailed model to a high degree of accuracy. J Neurophysiol 92(2):959–976

Jolivet R, Rauch A, Lüscher HR, Gerstner W (2006a) Integrate-and-Fire models with adaptation are good enough: predicting spike times under random current injection. Adv Neural Inform Process Syst 18:595–602

Jolivet R, Rauch A, Lüscher HR, Gerstner W (2006b) Predicting spike timing of neocortical pyramidal neurons by simple threshold models. J Comput Neurosci 21(1):35–49

Jolivet R, Kobayashi R, Rauch A, Naud R, Shinomoto S, Gerstner W (2008) A benchmark test for a quantitative assessment of simple neuron models. J Neurosci Meth 169(2):417–424

Keat J, Reinagel P, Reid RC, Meister M (2001) Predicting every spike: a model for the responses of visual neurons. Neuron 30:803–817

Kempter R, Gerstner W, van Hemmen JL, Wagner H (1998) Extracting oscillations: neuronal coincidence detection with noisy periodic spike input. Neural Comp 10:1987–2017

Keren N, Peled N, Korngreen A (2005) Constraining compartmental models using multiple voltage recordings and genetic algorithms. J Neurophysiol 94(6):3730–3742

Kistler WM, Gerstner W, van Hemmen JL (1997) Reduction of Hodgkin-Huxley equations to a single-variable threshold model. Neural Comp 9:1015–1045

Kobayashi R, Shinomoto S (2007) State space method for predicting the spike times of a neuron. Phys Rev E 75(1):11,925

La Camera G, Rauch A, Lüscher HR, Senn W, Fusi S (2004) Minimal models of adapted neuronal response to in vivo-like input currents. Neural Comp 16:2101–2124

Lansky P, Sanda P, He J (2006) The parameters of the stochastic leaky integrate-and-fire neuronal model. J Comput Neurosci 21(2):211–223

Larkum ME, Zhu JJ, Sakmann B (1999) A new cellular mechanism for coupling inputs arriving at different cortical layers. Nature 398(6725):338–341

Larkum ME, Senn W, Lüscher HR (2004) Top-down dendritic input increases the gain of layer 5 pyramidal neurons. Cereb Cortex 14:1059–1070

Leibold C, Gundlfinger A, Schmidt R, Thurley K, Schmitz D, Kempter R (2008) Temporal compression mediated by short-term synaptic plasticity. PNAS 105(11):4417

MacLeod K, Backer A, Laurent G (1998) Who reads temporal information contained across synchronized and oscillatory spike trains. Nature 395:693–698

Markram H (2006) The blue brain project. Nat Rev Neurosci 7(2):153–160

Marmarelis VZ, Berger TW (2005) General methodology for nonlinear modeling of neural systems with Poisson point-process inputs. Math Biosci 196(1):1–13

Mullowney P, Iyengar S (2008) Parameter estimation for a leaky integrate-and-fire neuronal model from ISI data. J Comput Neurosci 24:179–194

Paninski L, Pillow JW, Simoncelli EP (2005) Comparing integrate-and-fire models estimated using intracellular and extracellular data. Neurocomputing 65(66):379–385

Pillow JW, Paninski L, Uzzell VJ, Simoncelli EP, Chichilnisky EJ (2005) Prediction and decoding of retinal ganglion cell responses with a probabilistic spiking model. J Neurosci 25(47):11,003–11,013

Prinz AA, Billimoria CP, Marder E (2003) Alternative to hand-tuning conductance-based models: construction and analysis of databases of model neurons. J Neurophysiol 90(6):3998–4015

Prinz AA, Bucher D, Marder E (2004) Similar network activity from disparate circuit parameters. Nat Neurosci 7(12):1345–1352

Rauch A, La Camera G, Lüscher HR, Senn W, Fusi S (2003) Neocortical pyramidal cells respond as integrate-and-fire neurons to in vivo-like input currents. J Neurophysiol 90:1598–1612

van Rossum MCW (2001) A novel spike distance. Neural Comp 13:751–763

Schaefer AT, Larkum ME, Sakmann B, Roth A (2003) Coincidence detection in pyramidal neurons is tuned by their dendritic branching pattern. J Neurophysiol 89(6):3143–3154

Song D, Chan RH, Marmarelis VZ, Hampson RE, Deadwyler SA, Berger TW (2007) Nonlinear dynamic modeling of spike train transformations for hippocampal-cortical prostheses. IEEE Trans Biomed Eng 54:1053–1066

Vanier MC, Bower JM (1999) A comparative survey of automated parameter-search methods for compartmental neural models. J Comput Neurosci 7(2):149–171

Victor JD, Purpura KP (1996) Nature and precision of temporal coding in visual cortex: a metric-space analysis. J Neurophysiol 76:1310–1326

Victor JD, Purpura KP (1997) Metric-space analysis of spike trains: theory, algorithms and application. Netw Comp Neural Syst 8:127–164

Wang Y, Gupta A, Toledo-Rodriguez M, Wu CZ, Markram H (2002) Anatomical, physiological, molecular and circuit properties of nest basket cells in the developing somatosensory cortex. Cereb Cortex 12:395–410

Weaver CM, Wearne SL (2006) The role of action potential shape and parameter constraints in optimization of compartment models. Neurocomputing 69(10–12):1053–1057