

A benchmark test for a quantitative assessment of simple neuron models

Renaud Jolivet^{a,*}, Ryota Kobayashi^b, Alexander Rauch^c, Richard Naud^d,
Shigeru Shinomoto^b, Wulfram Gerstner^d

^a Center for Psychiatric Neuroscience, University of Lausanne, 1015 Lausanne, Switzerland

^b Department of Physics, Kyoto University, Kyoto 606-8502, Japan

^c Max-Planck Institute for Biological Cybernetics, 72076 Tübingen, Germany

^d Laboratory of Computational Neuroscience, Brain Mind Institute, Ecole Polytechnique Fédérale de Lausanne (EPFL), 1015 Lausanne, Switzerland

Received 8 September 2007; received in revised form 1 November 2007; accepted 10 November 2007

Abstract

Several methods and algorithms have recently been proposed that allow for the systematic evaluation of simple neuron models from intracellular or extracellular recordings. Models built in this way generate good quantitative predictions of the future activity of neurons under temporally structured current injection. It is, however, difficult to compare the advantages of various models and algorithms since each model is designed for a different set of data. Here, we report about one of the first attempts to establish a benchmark test that permits a systematic comparison of methods and performances in predicting the activity of rat cortical pyramidal neurons. We present early submissions to the benchmark test and discuss implications for the design of future tests and simple neurons models.

© 2007 Elsevier B.V. All rights reserved.

Keywords: Spike timing prediction; Integrate-and-Fire model; Spike Response Model; Benchmark test

1. Introduction

Neurons communicate by generating action potentials that are transmitted to other neurons in the network. Action potentials are generated in response to transmembrane currents elicited by presynaptic activation of various receptor types. Despite years of research, the exact nature of the neural code, that is how presynaptic activity is processed and encoded in outgoing action potentials, is still unknown. Is the neuronal firing rate sufficient to describe neural activity or does the timing of spikes on a millisecond timescale matter as well? Following the seminal work of Hodgkin and Huxley (1952), a lot of effort has been spent to build and study biophysically detailed models of single neuron electrical activity. These models can reproduce a large variety of neuronal behaviors as observed in experiments by a suitable combination of different ion currents (Bower and

Beeman, 1995). However, only few studies have focused on the methodology of designing algorithms for automatic fitting of such models to data, so as to arrive at models with a quantitative predictive power (Druckmann et al., 2007; Huys et al., 2006; Markram, 2006; Prinz et al., 2003, 2004).

In contrast to detailed Hodgkin–Huxley models, very simple models only have a small number of parameters which can be automatically and easily extracted from electrophysiological recordings. As early as the 1970s and 1980s, neuroscientists have tried to develop methods for the evaluation of simple neuron models from neural data. Brillinger and Segundo, in particular, have used maximum-likelihood and optimal filtering techniques to evaluate the linear response curve and firing probability of neuronal membranes thus laying down the foundations for more modern approaches (Brillinger, 1988a,b; Brillinger and Segundo, 1979). More recently, different groups tried to extract not only parameters of interest from data but also to build neuron models with a true quantitative predictive power. More specifically, 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). Beyond the output

* Corresponding author at: EPFL SV BMI LNDC, Station 15, CH-1015 Lausanne, Switzerland. Tel.: +41 21 693 1652; fax: +41 21 693 1650.

E-mail address: renaud.jolivet@a3.epfl.ch (R. Jolivet).

URL: <http://icwww.epfl.ch/~rjolivet> (R. Jolivet).

firing rate, Keat et al. have shown that the precise spike time 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 on retinal ganglion cells by Pillow et al. (2005). 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., 2006; Paninski et al., 2005) and in vivo (Lansky et al., 2006). Finally, alternative methods have been proposed that still wait to be tested on experimental recordings (Jolivet and Gerstner, 2004; Kobayashi and Shinomoto, 2007).

While simple models can, at least qualitatively, reproduce a broad range of observed neuronal behaviors (Izhikevich, 2004), their simplicity also permits to mathematically analyze questions of neural coding (Arcas et al., 2003; Brunel et al., 2003; Keat et al., 2001; Pillow et al., 2005). Developing efficient simple models with quantitative predictive power is also of importance for implementation in neural prostheses where such models could be simulated at low cost or built in silico [see e.g. (Marmarelis and Berger, 2005; Song et al., 2007)].

Despite this intense activity, the community still lacks a benchmark test that could be used as a reference to compare cost and performances of different methods. Here, we describe such a benchmark test and report the first results from an international competition¹. In short, the goal is to predict the spike times of a layer-5 pyramidal neuron recorded from the rat somatosensory cortex under current injection in various discharge regimes. A first set of spike trains was made publicly available together with the corresponding stimuli for model evaluation. Participants had to predict the spike times with a precision of ± 2 ms for a different dataset for which only the stimuli were provided. Note that this initiative differs from the recent *Neural Prediction Challenge*² in the sense that the goal is to design a model that predicts spike times in response to a fluctuating current 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. The benchmark test is described in the next section. A summary of results and submissions is then presented in section 3 and discussed in the last section.

2. Methods

2.1. Electrophysiological recordings

Data used for the challenge have been extensively described in Refs. (Jolivet et al., 2006; La Camera et al., 2006; Rauch et al., 2003) and we refer interested readers to these publications for details of the experimental protocol. In short, parasagittal slices of rat somatosensory cortex (300 μ m thick) were prepared from 15- to 40-day-old female and male Wistar rats according to the institutional guidelines. We recorded in current-clamp whole

cell configuration from the soma of layer 5 regular spiking pyramidal cells (McCormick et al., 1985). Four cells were recorded and trial repetitions of the input were performed ($N=4$ repetitions). The input was generated with an Ornstein–Uhlenbeck process (Tuckwell, 1988). The total injected current $I(t)$ is given by

$$I(t + dt) = I(t) - \frac{I(t)}{\tau_I} dt + m_I dt + s_I \xi(t) \sqrt{dt} \quad (1)$$

where m_I and s_I are parameters and $\xi(t)$ is a zero-mean, unit-variance Gaussian random variable, updated at every time step. The process was generated and injected at a rate of 5 kHz ($dt=0.2$ ms) and the correlation length τ_I was 1 ms. The resulting current $I(t)$ has a stationary Gaussian distribution with mean $\mu_I = m_I \tau_I$ and variance $\sigma_I^2 = s_I^2 \tau_I / 2$ (Cox and Miller, 1965). μ_I and σ_I were varied as follows: the total range $0 < \mu_I < 650$ pA and $0 < \sigma_I < 350$ pA was discretized and then explored in random order to prevent correlations over time. The duration of the stimulation was 6.8 s long for each pair of parameters μ_I, σ_I . The intervals between successive stimulations were 50–60 s long.

2.2. Measuring the similarity between two spike trains

In order to measure the similarity or dissimilarity between two spike trains and assess the quality of the predictions of simple models, we need a measure to compare spike trains as predicted by the model to spike trains as generated by the original cell. One possibility consists of comparing output firing rates (Hansel and Mato, 2003; Rauch et al., 2003). This is a very effective method but it misses all temporal structure in spike trains. Several 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; Kistler et al., 1997; MacLeod et al., 1998) or on cost functions (Aronov and Victor, 2004; van Rossum, 2001; Victor and Purpura, 1997, 1996). In precedent reports, some of us have consistently used the *coincidence factor* Γ as defined in Refs. (Clopath et al., 2007; Jolivet and Gerstner, 2004; Jolivet et al., 2004, 2006; Kistler et al., 1997). The coincidence factor can be computed quickly and easily. It is written

$$\Gamma_k = \frac{N_{\text{coinc}} - \langle N_{\text{coinc}} \rangle}{1/2(N_{\text{data}}^k + N_{\text{model}}) N} \quad (2)$$

where N_{data}^k is the number of spikes in the k th reference spike train S_{data} , N_{model} is the number of spikes in the predicted spike train S_{model} that is compared with the reference spike train, N_{coinc} is the number of coincidences with precision Δ between the two spike trains, and $\langle N_{\text{coinc}} \rangle = 2f\Delta N_{\text{data}}^k$ is the expected number of coincidences generated by a homogeneous Poisson process with the same rate f as the spike train S_{model} . The factor $N = [1 - 2f\Delta]$ normalizes Γ to a maximum value of one which is reached if and only if the spike train of the model reproduces exactly that of the cell. A homogeneous Poisson process with the same number of spikes as the minimal model would yield $\Gamma = 0$, which is, therefore, the chance level.

The sole free parameter in the estimation of Γ is the coincidence window $\pm\Delta$. Γ is relatively independent of the exact

¹ <http://icwww.epfl.ch/QuantNeuronMod2007/>.

² <http://neuralprediction.berkeley.edu/>.

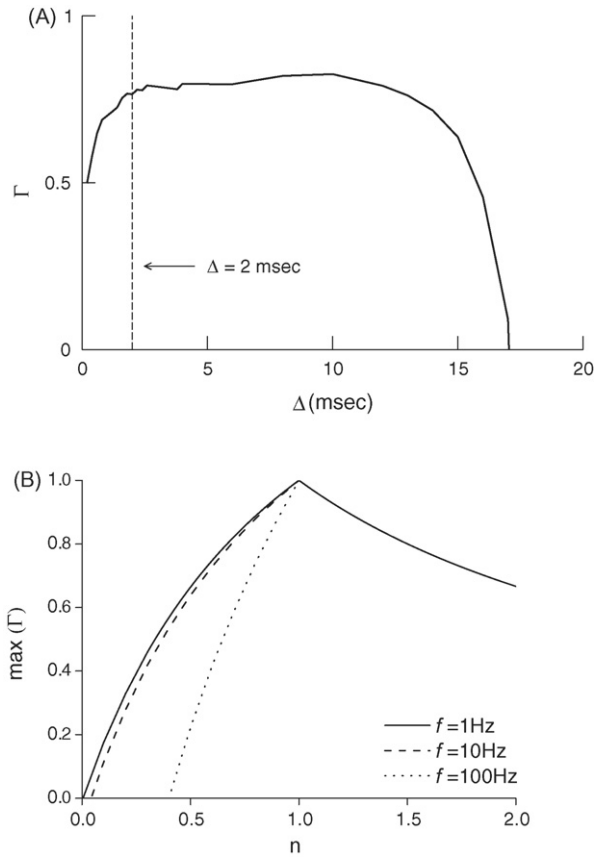


Fig. 1. Robustness and limitations of the coincidence factor Γ . (A) The coincidence factor Γ does not significantly depend on the width of the coincidence window Δ . $\Gamma(\Delta)$ is plotted for a comparison between a spike train from a layer 5 pyramidal neuron and the spike train predicted by a simple neuron model fitted with parameters to this specific neuron following the method presented in Ref. (Jolivet et al., 2006). In this specific example, the pyramidal neuron discharges at approximately 21 Hz and the model reaches a performance of $\Gamma = 0.77$ for $\Delta = 2$ ms. (B) The maximal theoretical value of Γ that one can reach with $\Delta = 2$ ms is plotted versus n defined as the ratio between the predicted number of spikes and the correct number of spikes $n = N_{\text{model}}/N_{\text{data}}$. For example, the value Γ for $n = 2$ was calculated from Eq. (3) by assuming that the model predicts all N_{data} spikes correctly, but places as many additional spikes at random locations in between the “correct” ones. The penalty in case of an incorrect prediction of the target frequency is asymmetric: predicting twice as many spikes as the real neuron gives a higher value of Γ than predicting only half of its spikes.

value of Δ . As illustrated in Fig. 1A, the value is approximately constant for values $2 \leq \Delta \leq 12$ ms. By convention, we chose $\Delta = 2$ ms which is about the temporal width of an action potential in cortical neurons. Therefore, two spikes are considered coincident while they overlap. The coincidence measure Γ is a relatively robust measure of spike train similarity (Jolivet and Gerstner, 2004; Jolivet et al., 2004; Jolivet et al., 2006; Kistler et al., 1997). It should be emphasized, however, that it imposes an asymmetric penalty when the predicted number of spikes differs from the true one (Fig. 1B). In particular, a high value of Γ (say $\Gamma = 0.8$) can be reached even if the predicted frequency significantly exceeds the target frequency. It is, therefore, necessary to control that the predicted frequency does not differ from the target frequency by a significant fraction (Clopath et al., 2007).

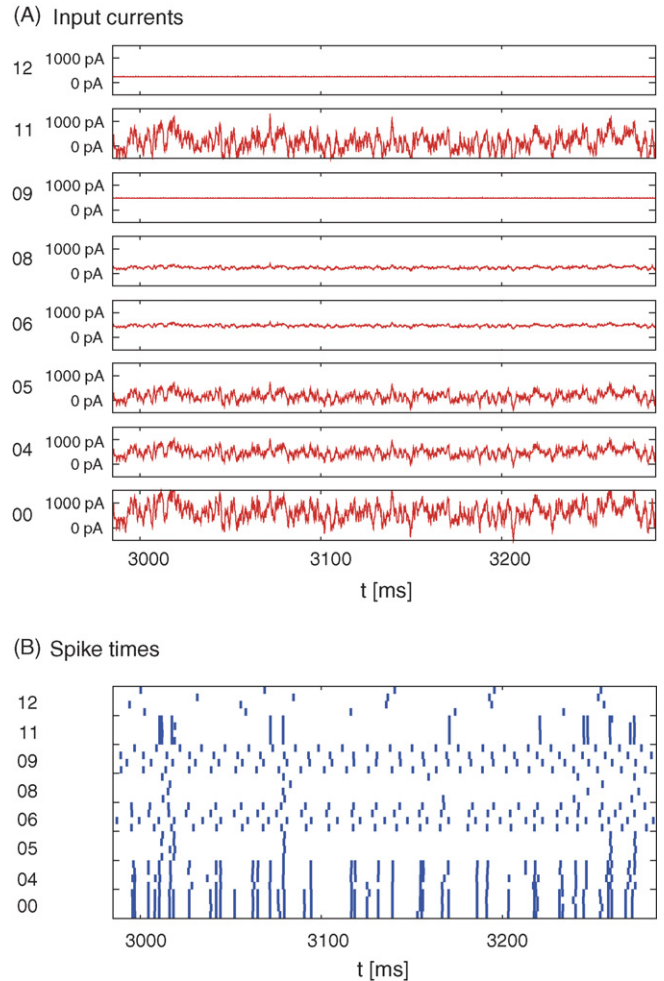


Fig. 2. (Top) Eight fluctuating currents injected to a target neuron. The y-axis ranges from -500 pA to 1500 pA. (Bottom) Spike sequences evoked by the fluctuating currents. Four trials were carried out for each identical current.

2.3. Benchmark test

As it is typical for this type of tests, the dataset that consists of all recordings for one pyramidal neuron was split in two parts with about two third being made available for model evaluation (“training set” containing stimuli and neuronal responses) and the remaining data being used for performance measure (“test set” containing the test stimuli only). In total, the training set contained eight different stimuli [current traces generated by Eq. (1), see Fig. 2A] and for each stimulus, four neuronal voltage traces (corresponding to four repetitions of the stimulus) were given. The test set consisted of other stimuli. To evaluate and rank the models, we use the global performance Γ_A defined as follows

$$\Gamma_A = \frac{1}{K} \sum_{k=1}^K \frac{\Gamma_k}{\hat{\Gamma}_k} \quad (3)$$

with Γ_k being measured between the target spike train and the predicted spike trains [see Eq. (2)] and the sum running over the K stimuli of the test dataset. $\hat{\Gamma}_k$ is the intrinsic reliability of the neuron. It is measured using Γ [Eq. (2)] evaluated not between a model and the experimental trains but between all possible pairs

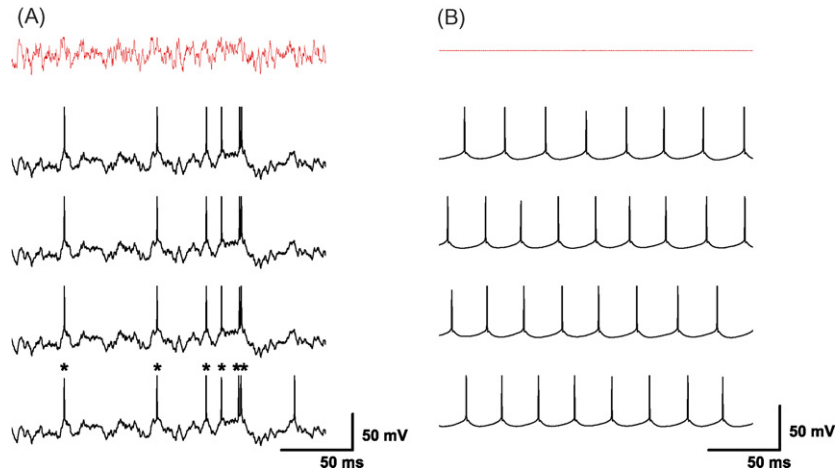


Fig. 3. Two regimes of activity of the pyramidal neuron. (A) A strongly fluctuating input (red trace) tends to generate very irregular but highly reliable spiking (black traces). In this example, all spikes are repeated in all four trial repetitions (asterisks) except one generated only once during the last repetition (unlabelled). (B) On the opposite, a constant or quasi-constant input tends to generate regular but unreliable spiking.

of the four trial repetitions corresponding to a specific stimulus. We find different discharge regimes, since neurons elicit action potentials with different reliability depending on the nature of input (Mainen and Sejnowski, 1995). Strongly fluctuating inputs tend to generate very irregular but highly reliable spiking. In contrast to this, constant or slow stimuli tend to generate regular but unreliable spiking (Fig. 3). It is, therefore, necessary to measure the spike time predictions of models relative to the intrinsic reproducibility of spikes by the neuron itself. In the reliable regime ($\hat{I}_k > 0.7$, Fig. 3A), we expect very good predictions of the spike times while in the unreliable regime (Fig. 3B), we only expect the overall firing frequency to be correct.

2.4. Model fitting methods

In the following, we present and compare four different and independent methods for spike time prediction, all used on the benchmark data mentioned above: (1) we estimated the membrane potential with the autoregressive (AR) model and predicted the spike occurrences using a dynamic threshold; (2) we simply carbon-copied (CC) and pasted the spike times of another experiment, based on our observation that fluctuating currents of similar temporal structures tend to evoke similar spike sequences; (3) we estimated the membrane potential with the Spike response model (SRM) and predicted spike times with a dynamic threshold and (4) we did the same with an adaptive exponential integrate-and-fire model (aEIF). A brief description of these four methods follows. There is room for other mathematical models for mimicking the membrane potential such as the Hodgkin–Huxley model (Hodgkin and Huxley, 1952; Tsubo et al., 2004) or the linear filter model (Westwick and Kearney, 2003) but we focused on these four.

2.4.1. The autoregressive model (AR)

In our first strategy, we introduced a mathematical model for mimicking training data that consist of both an input current and the membrane potential. The model adjusted to training data was applied to a novel test input current for estimating the

hidden membrane potential of the neuron. Then, the spike times were predicted by applying a dynamic threshold to the estimated membrane potential (Fig. 4).

We start with the autoregressive (AR) or the autoregressive exogenous (ARX) model and write

$$u(t + \Delta t) = Au(t) + BI(t) \quad (4)$$

where $u(t)$ is the membrane potential measured with respect to rest to be estimated and $I(t)$ is the input current sampled at an interval of Δt . Given a training data set, the model parameters A and B can be determined by solving the Yule–Walker equation (Chatfield, 2003). We expected the parameters A and B that characterize the integration mechanisms of a given neuron to be invariant with respect to various currents, but they were in practice dependent on the choice of input currents. We obtained an empirical rule that describes how the parameters A and B depend on the mean μ_I and variance σ_I^2 of the input current. In estimating the hidden potential for a test input current, we iterated the AR model, Eq. (4), whose

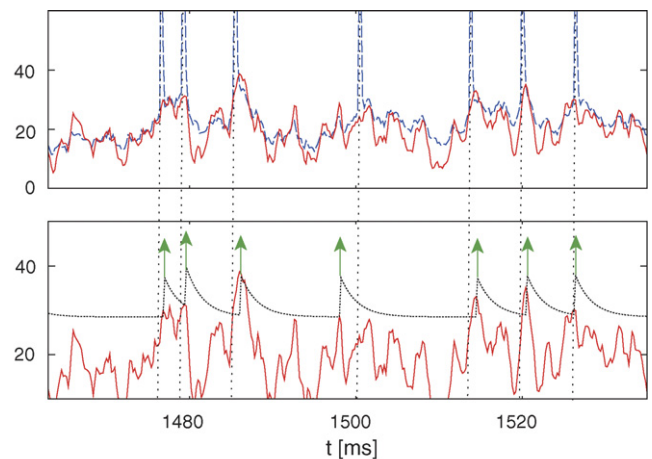


Fig. 4. (Top) Membrane potential of the biological neuron (dashed line) and the AR model (solid line). (Bottom) Spike time prediction according to the dynamic threshold assumption.

parameters A and B had been determined with the empirical rule.

In predicting the spike times from the estimated membrane potential, it would be worthwhile to implement the state space method that estimates the probability of spike occurrence, maximally utilizing the information of not only the voltage but also its time derivative (Kobayashi and Shinomoto, 2007). In this contribution, we simply used a dynamic threshold (Geisler and Goldberg, 1966; Jolivet et al., 2004) to the estimated membrane potential

$$\frac{d\vartheta}{dt} = -\frac{\vartheta - \vartheta_0}{\tau_\vartheta} \quad (5)$$

$$\vartheta \rightarrow \vartheta + A_\vartheta, \quad \text{when fired}$$

The model parameters ϑ_0 , A_ϑ and τ_ϑ were determined as follows. First, ϑ_0 was determined so that the model generates spikes with the rate of the hidden spikes that the test current would have been evoking. For this purpose, we constructed an empirical rule for estimating the spike rate from the mean and variance of an input current. Second, the other parameters α and τ_ϑ were determined so that the coincidence score Γ would be maximized, as empirically confirmed by cross-validating the training data.

2.4.2. The carbon-copy model (CC)

In our second strategy, we did not employ any mathematical model, but simply copied and pasted spike times of some training data set. In the experiments, fluctuating currents were generated with identical random numbers, and only their means and variances were varied. It can be observed from Fig. 2 that spikes evoked by different current amplitudes, but same time sequences coincided fairly often.

The similarity of the spike sequences greatly depends on the means and variances of the currents. Given a test input current, we wish to select a spike sequence from training data that would attain the highest coincidence score Γ . For this purpose, we invented a similarity metric measuring a sort of (asymmetrical) distance from a training current to a test, using their means and variances.

2.4.3. The spike response model (SRM)

As a third method, we have used the spike response model (Gerstner and Kistler, 2002). It is written

$$u(t) = \eta(t - \hat{t}) + \int_0^{+\infty} \kappa(s)I(t - s) ds \quad (6)$$

with u the membrane voltage of the neuron and I the external driving current. The kernel κ models the integrative properties of the membrane. The kernel η acts as a template for the shape of spikes (usually highly stereotyped). Like in the integrate-and-fire model, the model neuron fires each time that the membrane voltage u crosses the threshold ϑ from below

$$\text{if } u(t) \geq \vartheta(t) \text{ and } \frac{d}{dt}u(t) \geq \frac{d}{dt}\vartheta(t), \quad \text{then } \hat{t} = t \quad (7)$$

Here, the threshold includes a mechanism of spike-frequency adaptation. ϑ is given by Eq. (5) above. In other words, each time

that a spike is fired, the threshold ϑ is increased by a fixed amount A_ϑ . It then decays back to its resting value ϑ_0 with time constant τ_ϑ . During discharge at rate f , the threshold fluctuates around the average value

$$\bar{\vartheta} \approx \vartheta_0 + \alpha f \quad (8)$$

where $\alpha = A_\vartheta \tau_\vartheta$. This type of adaptation mechanism has been shown to constitute a universal model for spike-frequency adaptation (Benda and Herz, 2003). During the model estimation procedure, we use a classic constant threshold $\vartheta(t) = \vartheta_{\text{cte}}$ determined independently for each pair of parameters (μ_I, σ_I) . The result is then transformed in the adaptive threshold of Eq. (5) by the procedure detailed now.

Details of the mapping technique have been extensively described in Jolivet et al. (2006). In short, it is a systematic step-by-step evaluation and optimization procedure based on intracellular recordings. It consists in sequentially evaluating kernels (η and κ) and parameters [A_ϑ, ϑ_0 and τ_ϑ in Eq. (5)] that characterize a specific instance of the model. The consecutive steps of the procedure are as follows

1. Extract the kernel η from a sample voltage recording by spike triggered averaging. For the sake of simplicity, we assume that the mean drive vanishes, $\mu_I = 0$.
2. Subtract η from the voltage recording to isolate the subthreshold fluctuations.
3. Extract the kernel κ by a Wiener–Hopf optimal filtering technique (Jolivet et al., 2004; Wiener, 1958). This step involves a comparison between the subthreshold fluctuations and the corresponding input current.
4. Find the optimal constant threshold ϑ_{cte} for each pair of parameters (μ_I, σ_I) . The optimal value of ϑ_{cte} is the one that maximizes the coefficient Γ .
5. Plot ϑ_{cte} as a function of the firing frequency and run a linear regression on the results of step 4. ϑ_0 is identified with the value of the fit at $f=0$ and α with the slope [see Eq. (8)].
6. Optimize A_ϑ for the best performances (again measured with Γ), τ_ϑ is defined as $\tau_\vartheta = \alpha/A_\vartheta$.

2.4.4. The adaptive exponential integrate-and-fire model (aEIF)

We investigated the performance of a fourth method, the adaptive exponential integrate-and-fire (aEIF) model (Brette and Gerstner, 2005). It is formulated by a system of differential equations

$$C \frac{du}{dt} = -g_L(u - E_L) + g_L \Delta_T \exp \left[\frac{(u - V_T)}{\Delta_T} \right] + I - w \quad (9)$$

$$\tau_w \frac{dw}{dt} = a(u - E_L) - w \quad (10)$$

The model is said to spike when u reaches 20 mV; u and w are then reinitialized with the reset condition

$$\text{if } u \geq 20 \text{ mV}, \quad \text{then } u \rightarrow E_r \text{ and } w \rightarrow w + b \quad (11)$$

The state of the system is defined by the membrane potential u and the adaptive current w , also called the recovery variable.

In total, there are nine parameters: The parameters g_L , C and E_L regulate the integrative properties of the neuron, we refer to these as the leak conductance, the capacitance and the leak reversal potential, respectively. The parameters Δ_T and V_T define the spike initiation, τ_w describes the timescale of the recovery variable and ‘ a ’ its sensitivity to subthreshold activity. The last two parameters define the reset potential E_r and the spike-triggered adaptation b .

The aEIF is a dynamical system very similar to the model proposed by Izhikevich (2003) to the difference that in the aEIF, the voltage equation [Eq. (9)] contains an exponential (Fourcaud-Trocmé et al., 2003) rather than a quadratic nonlinearity. The model can be seen as a two-dimensional reduction of the spike initiation in Hodgkin–Huxley models. In particular, there is not a strict threshold but a smooth spike initiation zone. It has been shown previously that the aEIF reproduces up to 96% of the spike times of a Hodgkin–Huxley-type model (Brette and Gerstner, 2005). For more details regarding this model, we refer the reader to the work of Brette and Gerstner (2005).

To fit the nine parameters of the aEIF, we resorted to a black box optimization scheme that is described in the following steps. First, the parameters g_L and E_L were fixed since they do not influence the maximal performance attainable by the model: Shifting both E_L and V_T by the same amount is equivalent to a mere shift in the voltage scale. Similarly, increasing g_L and C by the same factor does not change the membrane time constant C/g_L . The remaining seven parameters were fit using a genetic algorithm inspired by the study of Vanier and Bower (1999). We required the algorithm to minimize $1 - \Gamma^*$, where Γ^* is the average coincidence factor [Eq. (2)] calculated on the first 2 s of the “training set” traces. We used the genetic algorithm implementation available in MATLAB (*The Mathworks*, Natick MA) with 150 generation of 100 individuals and a crossover ratio of 0.6, which takes roughly 2 h to terminate. The initial population was composed of four parameter sets that had been hand-tuned to reproduce qualitatively four major types

of spiking patterns, i.e., adapting, initially bursting, fast and slow tonic spiking. The best parameters after the optimization were: $C = 72$ pF, $\Delta_T = 0.006$ mV, $V_T = -38$ mV, $\tau_w = -25$ ms, $a = -0.5$ nS, $b = 36$ pA, $E_L = -60$ mV and $g_L = -13$ nS. These are parameters used on the test stimuli.

3. Results

Results of the challenge are summarized in Tables 1 and 2. We received four submissions in total. Three submissions were generated using a deterministic model and one was generated using a stochastic model (Anonymous 2). In addition, the challenge organizers provided two submissions as a reference ex post facto. The raw coefficient Γ was evaluated by averaging the values Γ_k obtained over the whole test set ($\Gamma = \sum_k \Gamma_k / K$). Obviously, this average Γ strongly depends on the composition of the test set. For certain experimental stimuli, the pyramidal neuron was more reliable than for others (Fig. 3). Therefore, we expect that a test set composed mostly of intrinsically reliable spike trains is “easier” for prediction by a model and yields a much better Γ than a test set composed of mostly unreliable spike trains. The global performance Γ_A was, therefore, derived by scaling the raw Γ with the intrinsic reliability of the neuron (see Section 2). This scaled performance measure was then used to rank submissions [see Eq. (3)]. The coefficients for intrinsic reliability in the test set were $\hat{\Gamma} = 0.22, 0.76, 0.85$ and 0.89 . Table 1 reports values computed with all four test spike trains while Table 2 reports values computed with only the intrinsically reliable test spike trains (i.e., dropping results obtained on the test spike train with $\hat{\Gamma} = 0.22$). Results obtained on the test set after the challenge using the methods described in sections 2.4.3 and 2.4.4 are indicated at the bottom of Tables 1 and 2 and serve as a reference. A comparison between a target spike train and the corresponding spike train predicted using the method described in Ref. (Jolivet et al., 2006) can be seen in Fig. 5. It shows that $\Gamma_A \cong 0.69$ corresponds to a very good quantitative prediction

Table 1
Results of the challenge

Submission	Group	Reference	Raw Γ	Performance (Γ_A)
Submission 1	Kyoto/Kobayashi (AR)	See Section 2.4.1	0.49	0.72
Submission 2	Kyoto/Shinomoto (CC)	See Section 2.4.2	0.54	0.66
Submission 3	Anonymous 1	–	0.39	0.55
Submission 4	Anonymous 2	–	0.26	0.32
Reference 1	Lausanne/Jolivet (SRM)	See Section 2.4.3	0.56	0.69
Reference 2	Lausanne/Naud (aEIF)	See Section 2.4.4	0.70	0.82

Table 2
Results of the challenge on intrinsically reliable data only

Submission	Group	Reference	Raw Γ	Performance (Γ_A)
Submission 1	Kyoto/Kobayashi (AR)	See Section 2.4.1	0.60	0.73
Submission 2	Kyoto/Shinomoto (CC)	See Section 2.4.2	0.71	0.84
Submission 3	Anonymous 1	–	0.48	0.56
Submission 4	Anonymous 2	–	0.33	0.40
Reference 1	Lausanne/Jolivet (SRM)	See Section 2.4.3	0.70	0.84
Reference 2	Lausanne/Naud (aEIF)	See Section 2.4.4	0.70	0.83

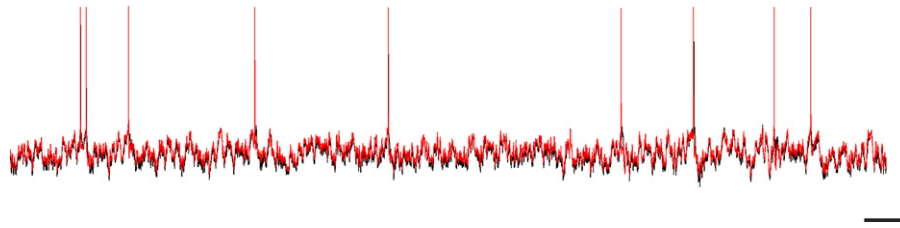


Fig. 5. Comparison between the spike train generated by a layer 5 pyramidal neuron (black trace) and the spike train predicted by a simple neuron model (red trace). See (Jolivet et al., 2006) for further details. Scale bars are 20 mV (vertical) and 100 ms (horizontal).

of not only the spike times but also the subthreshold fluctuations of the membrane voltage. Similarly, Fig. 4 shows a comparison between a target spike train and the corresponding spike train predicted using the AR method described in this paper.

Among the submissions to the challenge, the AR method yields the best performance on the complete test set beating the Spike Response Model (SRM) by about 5% but beaten by the adaptive exponential integrate-and-fire (aEIF) model by about 13% (Table 1). Considering a specific example, one observes that the AR model correctly predicts 729 spikes at the correct timing over a total of 795 spikes in the target spike train (93%) while the SRM predicts 657 spikes at the correct timing (83%) and the aEIF predicts 763 spikes at the correct timing (96%). The difference between the AR model and the SRM can be explained by the fact that in the AR model, the model parameters A and B were chosen independently for each stimulus whereas for the SRM and the aEIF model, a single parameter set was used across all stimuli. In this specific example, the corresponding raw Γ coefficients are $\Gamma = 0.597$ (AR), $\Gamma = 0.573$ (SRM) and $\Gamma = 0.571$ (aEIF). Since the measure Γ corrects for potential random coincidences, raw Γ coefficients are significantly lower than the percentage of predicted spikes. On the contrary, the SRM yields the best performance together with the carbon-copy (CC) technique and the aEIF model on the partial test set containing only intrinsically reliable spike sequences (Table 2). As one might expect from similar models, visual inspection confirms that the spikes correctly predicted by the SRM, by the aEIF and by the AR model are essentially the same.

4. Discussion

Despite an intense activity in trying to develop neuron models with a quantitative predictive power (Clopath et al., 2007; Huys et al., 2006; Jolivet et al., 2004, 2006; Keat et al., 2001; Kobayashi and Shinomoto, 2007; La Camera et al., 2004; Lansky et al., 2006; Paninski et al., 2005; Pillow et al., 2005; Prinz et al., 2003, 2004; Rauch et al., 2003), a standard test to compare the performances of one model versus the others is still lacking. Here, we described such a benchmark test for simple neuron models and reported the first results of a challenge on “Quantitative Neuron Modeling: Predicting every spike?” There were four submissions plus two reference models from earlier work (Clopath et al., 2007; Jolivet et al., 2006).

In this first round of the proposed benchmark test, the AR model described above and developed at Kyoto University was the best submission. It reached a global performance $\Gamma_A \cong 0.72$

on the test set meaning it was able to predict on average about 72% of spikes that one may expect to be predictable at the correct timing ± 2 ms. The AR model reached a slightly lower performance than the reference method using the aEIF model which reached $\Gamma_A \cong 0.82$. This result alone illustrates the capacity of simplified models in predicting the spike train of cortical pyramidal neurons while staying at an analytically tractable level. As we already observed in precedent studies, it is relatively easy to predict the subthreshold voltage but rather difficult to find a good criterion to decide when spikes are elicited (Jolivet et al., 2006; Kobayashi and Shinomoto, 2007). Since the global performance (Γ_A) reported in this paper is far below one, there is still room for improvement of such a threshold criterion. As suggested by the very good performance of the aEIF model, we expect that methods more sophisticated than a simple threshold crossing of the voltage will help improve performances of simple models (Kobayashi and Shinomoto, 2007) as well as fuel the debate on the neural code.

The fact that the carbon-copy method has achieved a fairly good performance has a significant implication: A neuron has a strong tendency of evoking similar spike sequences as long as the temporal structure of inputs is similar, even if their means and variances are very different (Fig. 2). We confirmed this fact with the numerical simulation of model neurons. In addition, neurons of different parameters have a tendency to generate coincident spikes in response to correlated inputs (Brette and Guigon, 2003). This fact is consistent with the idea of synfire chains (Abeles, 1991; Ikegaya et al., 2004). Indeed, it is likely that cortical neurons receiving correlated inputs evoke output spikes that coincide in time.

As a conclusion, this first attempt at setting a benchmark test illustrates the capabilities of simplified model neurons. It also shows that the best models at the moment have remarkably similar performances. In the future, we will extend the test by adding new data sets. In particular, we will add new recordings with injected currents or conductance injection using dynamic clamp generated with different random sequences to better sample the input space.

Acknowledgements

We thank Hans-Rudolf Lüscher for providing access to data. Felix Schürmann and Arnd Roth are thanked for commenting and helping with the organization of this benchmark test. AR was supported by a grant of the Swiss National Science Foundation (PBBSB-106816) and the M. and W. Lichtenstein Foundation (Switzerland). RN was supported by the Fonds Québécois de

Recherche sur la Nature et les Technologies (FQRNT). Modelling was partially supported by the European grant FACETS.

References

- Abeles M. *Corticonics*. Cambridge: Cambridge University Press; 1991.
- Arcas B, Fairhall A, Bialek W. Computation in a single neuron: Hodgkin and Huxley revisited. *Neural Comp* 2003;15:1715–49.
- Aronov D, Victor JD. Non-Euclidean properties of spike train metric spaces. *Phys Rev E* 2004;69:061905.
- Benda J, Herz AVM. A universal model for spike-frequency adaptation. *Neural Comp* 2003;15:2523–64.
- Bower J, Beeman D. *The book of genesis*. New York: Springer Verlag; 1995.
- Brette R, Gerstner W. Adaptive exponential integrate-and-fire model as an effective description of neuronal activity. *J Neurophysiol* 2005;94:3637–42.
- Brette R, Guigon E. Reliability of spike timing is a general property of spiking model neurons. *Neural Comp* 2003;15:279–308.
- Brillinger D. Maximum-likelihood analysis of spike trains of interacting nerve-cells. *Biol Cybern* 1988a;59:189–200.
- Brillinger D. The maximum likelihood approach to the identification of neuronal firing systems. *Ann Biomed Eng* 1988b;16:3–16.
- Brillinger D, Segundo J. Empirical examination of the threshold model of neuronal firing. *Biol Cybern* 1979;35:213–20.
- Brunel N, Hakim V, Richardson MJE. Firing-rate resonance in a generalized integrate-and-fire neuron with subthreshold resonance. *Phys Rev E* 2003;67:051916.
- Chatfield C. *The analysis of time series: an introduction*. 6th ed. London: Chapman and Hall/CRC; 2003.
- Clopath C, Jolivet R, Rauch A, Lüscher HR, Gerstner W. Predicting neuronal activity with simple models of the threshold type: adaptive exponential integrate-and-fire model with two compartments. *Neurocomp* 2007;70:1668–73.
- Cox D, Miller H. *The theory of stochastic processes*. New York: Chapman & Hall; 1965.
- Druckmann S, Banitt Y, Gidon A, Schürmann F, Markram H, Segev I. A novel multiple objective optimization framework for constraining conductance-based neuron models by experimental data. *Front Neurosci* 2007;1:7–18.
- Fourcaud-Trocme N, Hansel D, van Vreeswijk C, Brunel N. How spike generation mechanisms determine the neuronal response to fluctuating inputs. *J Neurosci* 2003;23:11628–40.
- Geisler CD, Goldberg JM. A stochastic model of repetitive activity of neurons. *Biophys J* 1966;6:53–69.
- Geisler WS, Albrecht DG, Salvi RJ, Saunders SS. Discrimination performance of single neurons—rate and temporal-pattern information. *J Neurophysiol* 1991;66:334–62.
- Gerstner W, Kistler W. *Spiking neurons models: single neurons, populations, plasticity*. Cambridge: Cambridge University Press; 2002.
- Hansel D, Mato G. Asynchronous states and the emergence of synchrony in large networks of interacting excitatory and inhibitory neurons. *Neural Comp* 2003;15:1–56.
- Hodgkin AL, Huxley AF. A quantitative description of membrane current and its application to conduction and excitation in nerve. *J Physiol-Lond* 1952;117:500–44.
- Huys QJM, Ahrens MB, Paninski L. Efficient estimation of detailed single-neuron models. *J Neurophysiol* 2006;96:872–90.
- Ikegaya Y, Aaron G, Cossart R, Aronov D, Lampl I, Ferster D, et al. Synfire chains and cortical songs: temporal modules of cortical activity. *Science* 2004;304:559–64.
- Izhikevich E. Simple model of spiking neurons. *IEEE Trans Neural Net* 2003;14:1569–72.
- Izhikevich E. Which model to use for cortical spiking neurons? *IEEE Trans Neural Net* 2004;15:1063–70.
- Jolivet R, Gerstner W. Predicting spike times of a detailed conductance-based neuron model driven by stochastic spike arrival. *J Physiol-Paris* 2004;98:442–51.
- Jolivet R, Lewis TJ, Gerstner W. Generalized integrate-and-fire models of neuronal activity approximate spike trains of a detailed model to a high degree of accuracy. *J Neurophysiol* 2004;92:959–76.
- Jolivet R, Rauch A, Lüscher HR, Gerstner W. Predicting spike timing of neocortical pyramidal neurons by simple threshold models. *J Comp Neurosci* 2006;21:35–49.
- Keat J, Reinagel P, Reid R, Meister M. Predicting every spike: a model for the responses of visual neurons. *Neuron* 2001;30:803–17.
- Kistler W, Gerstner W, van Hemmen J. Reduction of Hodgkin–Huxley equations to a single-variable threshold model. *Neural Comp* 1997;9:1015–45.
- Kobayashi R, Shinomoto S. State space method for predicting the spike times of a neuron. *Phys Rev E* 2007;75:011925.
- La Camera G, Rauch A, Lüscher HR, Senn W, Fusi S. Minimal models of adapted neuronal response to in vivo-like input currents. *Neural Comp* 2004;16:2101–24.
- La Camera G, Rauch A, Thurbon D, Lüscher H, Senn W, Fusi S. Multiple time scales of temporal response in pyramidal and fast spiking cortical neurons. *J Neurophysiol* 2006;96:3448–64.
- Lansky P, Sanda P, He JF. The parameters of the stochastic leaky integrate-and-fire neuronal model. *J Comp Neurosci* 2006;21:211–23.
- MacLeod K, Backer A, Laurent G. Who reads temporal information contained across synchronized and oscillatory spike trains? *Nature* 1998;395:693–8.
- Mainen Z, Sejnowski T. Reliability of spike timing in neocortical neurons. *Science* 1995;268:1503–6.
- Markram H. The blue brain project. *Nat Rev Neurosci* 2006;7:153–60.
- Marmarelis VZ, Berger TW. General methodology for nonlinear modeling of neural systems with Poisson point-process inputs. *Math Biosci* 2005;196:1–13.
- McCormick D, Connors B, Lighthall J, Prince D. Comparative electrophysiology of pyramidal and sparsely stellate neurons of the neocortex. *J Neurophysiol* 1985;54:782–806.
- Paninski L, Pillow J, Simoncelli E. Comparing integrate-and-fire models estimated using intracellular and extracellular data. *Neurocomp* 2005;65–66:379–85.
- Pillow JW, Paninski L, Uzzell VJ, Simoncelli EP, Chichilnisky EJ. Prediction and decoding of retinal ganglion cell responses with a probabilistic spiking model. *J Neurosci* 2005;25:11003–13.
- Prinz AA, Billimoria CP, Marder E. Alternative to hand-tuning conductance-based models: construction and analysis of databases of model neurons. *J Neurophysiol* 2003;90:3998–4015.
- Prinz AA, Bucher D, Marder E. Similar network activity from disparate circuit parameters. *Nat Neurosci* 2004;7:1345–52.
- Rauch A, La Camera G, Lüscher HR, Senn W, Fusi S. Neocortical pyramidal cells respond as integrate-and-fire neurons to in vivo-like input currents. *J Neurophysiol* 2003;90:1598–612.
- Song D, Chan RH, Marmarelis VZ, Hampson RE, Deadwyler SA, Berger TW. Nonlinear dynamic modeling of spike train transformations for hippocampal-cortical prostheses. *IEEE Trans Biomed Eng* 2007;54:1053–66.
- Tsubo Y, Kaneko T, Shinomoto S. Predicting spike timings of current-injected neurons. *Neural Netw* 2004;17:165–73.
- Tuckwell H. *Introduction to theoretic neurobiology*. Cambridge: Cambridge University Press; 1988.
- van Rossum M. A novel spike distance. *Neural Comp* 2001;13:751–63.
- Vanier MC, Bower JM. A comparative survey of automated parameter-search methods for compartmental neural models. *J Comp Neurosci* 1999;7:149–71.
- Victor JD, Purpura KP. Metric-space analysis of spike trains: theory, algorithms and application. *Netw-Comput Neural Syst* 1997;8:127–64.
- Victor JD, Purpura KP. Nature and precision of temporal coding in visual cortex: a metric-space analysis. *J Neurophysiol* 1996;76:1310–26.
- Westwick DT, Kearney RE. *Identification of nonlinear physiological systems*. Piscataway: Wiley-IEEE Press; 2003.
- Wiener N. *Nonlinear problems in random theory*. Cambridge: MIT Press; 1958.