

# Modeling cancellation of periodic inputs with burst-STDP and feedback

K. Bol<sup>1</sup>, G. Marsat<sup>2</sup>, J. F. Mejias<sup>1</sup>, L. Maler<sup>2,3</sup>, and A. Longtin<sup>1,3</sup>

<sup>1</sup>Dept. Physics, University of Ottawa, K1N 6N5 Ottawa, Canada.

<sup>2</sup>Dept. Cellular and Molecular Medicine, University of Ottawa, K1H 8M5 Ottawa, Canada.

<sup>3</sup>Centre for Neural Dynamics, K1H 8M5 Ottawa, Canada.

---

## Abstract

Prediction and cancellation of redundant information is an important feature that many neural systems must display in order to efficiently code external signals. We develop an analytic framework for such cancellation in sensory neurons produced by a cerebellar-like structure in wave-type electric fish. Our biologically plausible mechanism is motivated by experimental evidence of cancellation of periodic input arising from the proximity of conspecifics as well as tail motion. This mechanism involves elements present in a wide range of systems: 1) stimulus-driven feedback to the neurons acting as detectors, 2) a large variety of temporal delays in the pathways transmitting such feedback, responsible for producing frequency channels, and 3) burst-induced long-term plasticity. The bursting arises from back-propagating action potentials. Bursting events drive the input frequency-dependent learning rule, which in turn affects the feedback input and thus the burst rate. We show how the mean firing rate and the rate of production of 2- and 4-spike bursts (the main learning events) can be estimated analytically for a leaky integrate-and-fire model driven by (slow) sinusoidal, back-propagating and feedback inputs as well as rectified filtered noise. The effect of bursts on the average synaptic strength is also derived. Our results shed light on why bursts rather than single spikes can drive learning in such networks "online", i.e. in the absence of a correlative discharge. Phase locked spiking in frequency specific channels together with a frequency-dependent STDP window size regulate burst probability and duration self-consistently to implement cancellation.

*Keywords:* signal cancellation, long-term plasticity, bursting, cerebellum

---

## 1. Introduction

Many neural systems can respond to novel stimuli while filtering out redundant inputs. One of the most powerful illustrations of this skill is seen in the sensory "cocktail party effect", in which extraneous signals in an environment are identified and attenuated (Haykin & Chen, 2005). This is a powerful form of tunable noise suppression and likely involves the use of an adaptive filter (Roberts & Portfors, 2008; Sawtell & Williams, 2008), but the core mechanisms involved are far from known. Fundamentally, one form of the neural cocktail

party problem involves canceling arbitrary periodic signals in order to be sensitive to potentially weaker and novel stimuli. Filtering of predictable input is an example of redundancy reduction, thought to be a guiding principle in biological systems (Barlow, 2001). It is also known as subtraction of expectations (Gerstner & Kistler, 2002). Furthermore, cancellation of specific, identifiable inputs can be a partial solution to the blind source separation problem, e.g. using multiple recordings from the same source (Vorobyov et al., 2001). However, it remains unclear how actual neural circuits may perform such a task.

Consider a neuron in a network that receives both redundant sinusoidal input from sensory receptors and random novel stimuli. Let us assume that the identification of the redundant signal has already been accomplished; the task of the network is simply to eliminate it. What is the appropriate network to eliminate this signal? To cancel only predictable signals and not novel stimuli, this neuron must also receive input proportional to the neuron's past activity. This second input can be accomplished using a feedback pathway that encodes the activity of the neuron, or neurons entrained to the same sinusoidal input that do not cancel it. Optimally, this feedback would be adaptive, as slowly learning feedback would cancel the persistent sinusoid while leaving the novel stimuli unaffected. Adaptive feedback also prevents novel stimuli from affecting both the neuron and its feedback, thereby reducing echos in the network. Further, many feedback pathways, each with a unique phase delay but all entrained to the feedforward periodic signal, together create a delay line structure. Coupled with a suitable learning rule to alter their synapses, this network could cancel any redundant signal while leaving novel inputs intact.

Unfortunately, effective cancellation requires a fixed phase relationship between the feedback delay lines and the input, but it is the *time* delay of each delay line that is fixed anatomically. In order to maintain fixed phase delays, the feedback must be segregated into independent frequency channels. Otherwise, the system would have to relearn the appropriate synaptic strengths each time the input frequency changes. With a feedback composed of delay lines, independent frequency channels, and plastic synapses between the feedback and the intermediate neuron, the network can hypothetically cancel any periodic input frequency to reveal weak but potentially important inputs.

Such a network is thought to exist in the electroreceptor lateral line lobe of the weakly electric fish *Apteronotus leptorhynchus* (Bastian et al. (2004); Bol et al. (2011); see also Marsat et al. (2012) for a review). These fish continuously emit a high-frequency (600-1000 Hz) sinusoidal electric organ discharge (EOD) into their environment to sense their surroundings and communicate with conspecifics. The circuitry of their sensory neural network beautifully implements a sparse coding scheme, where specific cells respond to specific inputs (see Chacron et al. (2011) for a recent review). Small objects in the environment such as prey will create spatially localized amplitude modulations (AMs) of the EOD, whereas repetitive tail-bending or communication signals that arise from the beats of EODs from neighboring fish will induce spatially global AMs (Nelson & Maciver, 1999; Chen et al., 2005; Babineau et al., 2006). These almost sinusoidal AMs (or SAMs for short) are detected by electroreceptor afferents that densely cover the body of the fish (Carr & Maler, 1986), and linearly encode these signals into their firing rate modulation (Gussin et al., 2007; Chacron et al., 2000).

Electroreceptors provide feedforward input to deep pyramidal cells (DP) in

the electrosensory lateral line (ELL), the first electrosensory processing structure in the brain (Saunders & Bastian, 1984; Berman & Maler, 1999). These pyramidal cells then project to higher brain centers. Interestingly, it is known that a subpopulation of pyramidal cells, called superficial pyramidal (SP) cells, remove predictable global signals from their input to maximize detection of novel local stimuli (i.e. prey) (Bastian et al., 2004). This is putatively achieved thanks to a feedback pathway composed of delay lines - segregated into frequency channels - to destructively interfere with the global stimulus. Recently, the synaptic plasticity that shapes the feedback was found to be a novel correlative burst-timing dependent learning rule (Harvey-Girard et al., 2010).

In a recent study, a biophysically realistic model of the ELL architecture was built that effectively reproduced *in vivo* cancellation of global (i.e. full body) SAMs (Bol et al., 2011). That experimental and computational study also demonstrated the existence of frequency channels, where learning for the cancellation of one frequency did not affect that at other frequencies. It also built on the experimental and modeling work in Bastian et al. (2004), by including the novel burst-time-dependent learning rule uncovered by Harvey-Girard et al. (2010). This also means that the pyramidal model itself had to be more complex than integrate-and-fire (used in Bastian et al. (2004)), as it had to produce bursting.

In the present work, we study from a theoretical and computational point of view the cancellation of periodic signals by ELL superficial cells observed in our recent study. In particular, the stable distribution of synaptic strengths modified by a timing dependent learning rule is analytically predicted. This, in turn, leads to an estimation of the firing and burst rates as a function of the input. Therefore, the present work extends and clarifies the study of the cancellation mechanism observed and modelled in Bol et al. (2011) by providing a theoretical framework which may be used to better understand the cancellation circuits of the electric fish as well as other similar systems. Unlike other approaches that solve the stable distribution of synaptic strengths for random inputs (see e.g. Kepecs et al. (2002)), feedback delay lines allow the relative timing of pre-synaptic learning events to be deterministically predicted. Thus, the model may be written as a set of equations that can be solved self-consistently. The approach outlined here can be applied to any neural structure with delay lines and synaptic plasticity being used to shape a periodic input. It follows other recent efforts to explain the activity in neural networks with recurrent connections and STDP (see e.g. Gilson et al. (2009)).

Work on mormyrid weakly electric fish began to illuminate putative cancellation mechanisms fifteen years ago (Bell et al. (2000); see also Requarth & Sawtell (2011) for a recent review). This species emits electric pulses to survey changes in impedance in its environment caused by objects and other fish. The effect of these pulses on the environment is picked up by electroreceptors on the skin. But the pacemaker sending out the pulses to the electric organ also conveys spike discharges internally to ganglion neurons, to which the electroreceptors project. Through what has now become known as anti-Hebbian spike-time-dependent plasticity, these ganglions use this internal timing information (correlative discharge) to null out the redundant responses from the electroreceptors caused by the fish's own pulses, thus enabling the highlighting of novel stimuli (Roberts & Bell, 2000). The situation we confront here differs from this one in two main respects. Firstly, we consider wave-type weakly

electric fish, for which the redundant stimuli are ongoing rather than pulsatile. Secondly, there is no internal feedback in these animals, so the system must cancel redundant stimuli using only feedforward information. As we will see and analyze mathematically, this is achieved by using two classes of pyramidal neurons mentioned above (DP and SP) (the equivalent of the ganglion neurons in mormyrids) which are both in receipt of sensory input, but with one (DP) projecting feedback connections onto plastic synapses of the other (SP) via the EGp, a cerebellar-like structure (Bastian et al., 2004).

In Section 2, the weakly electric fish model, its assumptions and the underlying experimental data are introduced. In Section 3, the relationship between the input current and the firing rate of the model without feedback (i.e. feedforward only) is theoretically estimated. In addition, the effect of low-pass filtered Gaussian noise, rectified inputs and positive internal feedback (from backpropagation dynamics) on the mean firing and bursting rate of a leaky integrate-and-fire neural models is analyzed. In Section 4, external feedback from another cell population is introduced into the model. The firing rate is converted to an event learning rate, and the effect of a learning event on the average synaptic strength is derived. This method is also expanded to include multiple learning events. Finally, the average synaptic strength can be treated as an input and the resulting system of equations can be solved to generate the analytical estimates of the firing and burst rates of the model. In addition, an adiabatic approximation (i.e. separation of time scales) can be used to apply this approach to slowly varying inputs. This analysis extends STDP to include bursts and illuminates some of the characteristics of realistic adaptive filtering networks. Our results also shed light on why bursts and not single spikes were chosen as the appropriate learning event in this network.

## 2. Model

Our approach to the real neural system of interest is based on a leaky integrate-and-fire (LIF) neuron model with feedforward input and a series of feedback inputs that have modifiable synapses. This model was constrained to be biologically realistic, although the methods followed along this sections may be easily applicable to other systems. A simplified scheme of our system is shown in Figure 1.

### 2.1. Superficial cells

Superficial (SP) cell firing activity is modeled using the leaky integrate-and-fire formalism. The voltage of the SP cell evolves according to

$$\tau_m \frac{dV}{dt} = -V + [I + \sigma \xi_L(t) + \kappa \sin(2\pi ft)] + DAP(t) + \Lambda (w_s - gV) . \quad (1)$$

When the membrane potential,  $V$ , crosses the threshold,  $V_{thresh}$ , a spike is recorded and  $V$  is reset to  $V_{reset}$ . After that,  $V$  is maintained at  $V_{reset}$  for an absolute refractory period,  $\tau_{ref}$ , after which  $V$  continues to evolve according to Equation 1. Electroreceptor input is modeled, following a diffusion approximation due to the convergence of a large number of afferents (Gussin et al., 2007), as a bias current  $I$  that represents the mean excitatory bias of the input, plus low-pass filtered Gaussian noise,  $\xi_L(t)$  as in Doiron et al. (2003). Since

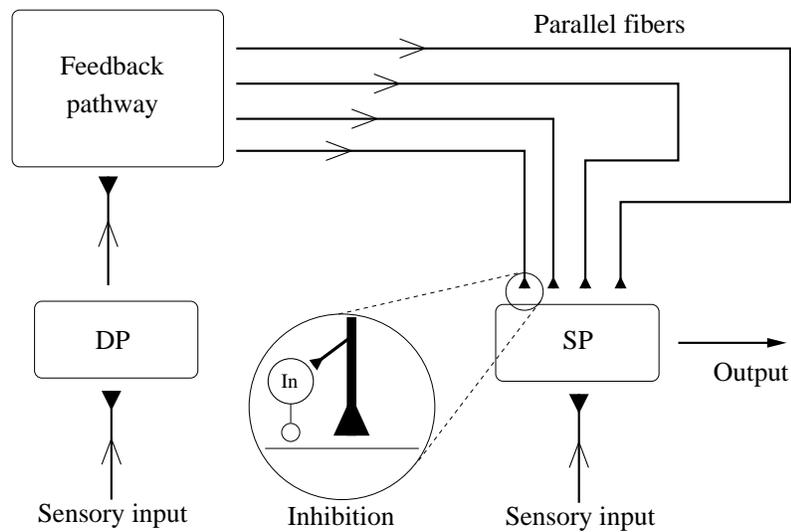


Figure 1: Simplified scheme of the neural system under study. Sensory input arrives to both superficial pyramidal (SP) neurons and deep pyramidal (DP) neurons. When this stimulus is global (that is, when it drives receptors all along the body of the animal), DP cells forward the input to higher structures, thus activating the feedback pathway that includes the nucleus Np as well as the cerebellar-like structure EGp. This pathway excites SP cells via parallel fibers. Each parallel fiber adds a unique temporal delay to the signal that arrives to SP cells from the feedback pathway, and also incorporates some degree of inhibition at the synaptic level via inhibitory interneurons (In). We will analyze how, after integrating both feedforward and feedback inputs, SP cells are able to cancel out the redundant components of the stimulus, allowing the transmission of novel information to other higher brain structures.

Parameter	Value	Parameter	Value	Parameter	Value
$V_{thresh}$	1	$\sigma$	0.759	$\eta_4$	$3.6 \times 10^{-3}$
$V_{reset}$	0	$f_{cut}$	500 Hz	$\eta_2$	$1.8 \times 10^{-3}$
$\tau_m$	7 ms	$\kappa$	0.39	$L_{w_4}$	100 ms
$\tau_{ref}$	0.7 ms	$\tau_w$	4900 s	$L_{w_2}$	10 ms
$I$	0.576	$w_{max}$	1	$g$	0.87

Table 1: Parameter values for Equations 1, 6, and 7 of the model to approximate experimental data. During analytical exploration of the model, parameters will be adjusted around these values.

electroreceptors linearly encode the stimulus (Gussin et al., 2007; Nelson et al., 1997; Chacron et al., 2000), the input of a sinusoidal AM with frequency  $f$  was modeled as  $\kappa \sin(2\pi ft)$ . In addition, as electroreceptor input is dominantly excitatory (GABA input attenuating AMPA and NMDA input (Berman & Maler, 1999)), the modeled feedforward input is rectified, and  $[\dots]$  in Equation 1 symbolizes rectification (that is,  $[x] = x$  if  $x > 0$ , and  $[x] = 0$  otherwise). This helped replicate the rectification observed in the pyramidal cell activity in response to the SAM (see Figure 2).

Low-pass filtered noise was generated by filtering Gaussian white noise through a fourth order Butterworth filter. After filtering, the noise is renormalized to have unit variance. The cutoff frequency of the filter is  $f_{cut}$  and the variance of the noise term is  $\sigma^2$ . The parameter  $\tau_m$  is the membrane time constant of the SP cell. During local stimulation the strength of the feedback,  $\Lambda$ , is set to zero as local stimulation does not drive the feedback. Other terms in Equation 1 are explained below. Parameter values for the model are summarized in Table 1.

## 2.2. Bursting dynamics

Superficial cell bursting both drives feedback plasticity at synapses at its apical dendrites, and has also been implicated in increased information transfer to downstream neurons (Oswald et al., 2004). The term  $DAP(t)$  in Equation 1 represents the depolarizing after-potential (DAP), an injection of current into the soma of the neuron after an action potential is fired due to presence of active channels in the cell’s dendrites. This effect has been modeled previously in superficial cells (Doiron et al., 2001; Noonan et al., 2003). Their model was used in this paper with minimal parameter changes:  $A$  and  $\gamma$  have simply been increased to match the bursting behavior observed *in vivo*.

Bursting arises from the following sequence of events. After the cell fires ( $V = V_{thresh}$ ) at time  $t_n$ , it will receive a DAP, i.e. a small current injection a short time later. This extra stimulation is modeled as a difference in alpha functions  $s(t, a)$  (Equation 3): one generated by the soma voltage, and the other by some mean dendrite voltage. If, however, the interval between this spike time  $t_n$  and the previous spike time  $t_{n-1}$  is less than the refractory period of the dendrite,  $r_d$ , then the DAP is inactive for the current spike. Such refractory period  $r_d$  is modeled as a dynamic variable  $r_d(t)$  that changes according to a secondary variable,  $b$ , which also controls the width of dendritic alpha function and updates whenever the neuron fires a spike.  $t_n^+$  refers to the time just after the most recent spike was fired. The equations governing the DAP (Noonan et al., 2003) are

Parameter	Value	Parameter	Value	Parameter	Value
$A$	0.6*	$B$	2	$\alpha$	20
$\gamma$	0.2*	$\beta$	0.35	$\tau$	1
$\tau_{ref}$	0.1	$D$	0.1	$E$	3.5

Table 2: Parameters used in the DAP model. Symbol (\*) indicates changes from Noonan et al. (2003). Time is in units of  $\tau_m = 7$  ms.

$$DAP(t) = \begin{cases} 0 & \text{if } t - t_n < r_s \\ \alpha \{s(t - t_n, \beta b(t_n^+)) - s(t - t_n, \gamma)\} & \text{if } t - t_n > r_s \text{ and } t_n - t_{n-1} > r_d(t_n^+) \\ 0 & \text{if } t - t_n > r_s \text{ and } t_n - t_{n-1} < r_d(t_n^+) \end{cases} \quad (2)$$

$$s(t, a) = \frac{te^{-\frac{t}{a}}}{a} \quad (3)$$

$$r_d(t) = D + Eb(t) \quad (4)$$

$$\frac{db}{dt} = -b/\tau + (A + Bb^2) \sum_n \delta(t - t_n) \quad (5)$$

The parameters used in the above equations are listed in Table 2.

### 2.3. Parallel fiber inputs

The feedback pathway is initiated by another population of pyramidal cells, called deep cells (DP), that do not exhibit the global cancellation response since they receive no feedback. They provide a reference point about the input to the canceling cells (SP). Via neurons in the nucleus praeminentialis (Np), deep cells drive granule cells in the cerebellar-like posterior eminentia granularis (EGp) at the same frequency as the global stimulus. These granule cells project massive numbers of excitatory parallel fibers (PFs) back into the ELL, which synapse onto SP cells as well as disynaptic inhibitory cells.

Due to difficulties in recording them, the firing activity of the granule cells in electric fish is unknown. Nevertheless, *in vivo* studies have shown that similar granule cells in mammals tend to burst to natural sensory input (Chadderton et al., 2004; Rancz et al., 2007) and phase-lock their bursting to sinusoidal input (D'Angelo et al. (2001), see also the Discussion section). Therefore, it was assumed that the total activity of each parallel fiber is one burst per stimulus period.

Nevertheless, input from different PFs will not reach SP cells simultaneously because of the parallel fiber network architecture. Each PF is unmyelinated and traces out a unique distance from its granule cell to each superficial cell. Depending on the spatial location of each cell in its respective structure as well as the relative location of the structures themselves (e.g. ipsilateral or contralateral) (Sas & Maler, 1987), the PFs that synapse onto any one SP cell will have a distribution of lengths. In addition, granule cells will not be active simultaneously because of delays in their input from different path lengths

between deep cells and granule cells (Carr & Maler, 1986) as well as noise and heterogeneities in the granule cell population.

Considering this distribution of delays, the total bursting PF input was assumed to be continuous in time. In the model, the feedback cycle associated with the input SAM cycle was discretized into 2.5 ms segments. This means that the number of segments,  $n_s(f)$ , changes with AM frequency (e.g. 100 feedback segments for a 4 Hz stimulus and 50 segments for an 8 Hz stimulus). Each segment, labeled  $s$ , becomes active at time  $t_s$ , has an inherent strength  $\Lambda$  and a synaptic weight  $w_s$  (see Equation 1), and then inactivates at  $t_s + 2.5$  ms. The total feedback input is a step-wise continuous and periodic signal composed of  $\Lambda$  multiplied by the synaptic strength,  $w_s$ , for each segment as time moves from segment to segment during a period. All weights were initialized at  $w_{max}$ .

A stable phase relationship for each segment and, hence, each weight, is necessary for cancellation. However, it is created from fixed temporal delays, and changing the AM frequency would require the weight distribution to relearn the appropriate values. Furthermore, for a fixed set of delays, parallel fiber activity would overlap at higher frequencies, and may not sufficiently cover the period of lower frequencies. For simplicity, it was assumed that each frequency is canceled independently and has its own unique synaptic weights for its collection of segments. Experimental evidence has corroborated this assumption (Bol et al., 2011).

The disynaptic inhibition to the SP cell induced by PFs is modeled as an extra shunting conductance,  $-gV$  in Equation 1, which is also multiplied by the PF feedback strength  $\Lambda$ . Since numerous PFs synapse onto one inhibitory cell, the sum of their input would be approximately constant. In addition, there is no experimental evidence that LTD occurs at these synapses (either on the input or output of the inhibitory interneurons). Therefore, disynaptic inhibition is fixed across phase and frequency.

#### 2.4. Burst definition and learning rule

Consistent with the definition of a burst that induces plasticity (Harvey-Girard et al., 2010), the model SP spike train was constantly analyzed for small (2 spikes within 15 ms) and large (4 spikes within 45 ms) bursts. These definitions of burst are only adopted here to simplify the theoretical calculations, as the quantitative behavior of our model does not depend sensitively on such assumptions, or even on the presence of strong intrinsic mechanisms for bursting generation (see Discussion for details). Note that spikes in each burst must be independent (i.e. there cannot be a small burst in a large burst, or a large and a small burst in 5 spikes). Since each parallel fiber segment produces a presynaptic burst arriving at the apical dendrite, there is one parallel fiber burst at every time  $t_s$  in the model, and thus PF bursts are spaced 2.5 ms apart. When the SP cell bursts under global stimulation at time  $t_B$ , the resulting volley of spikes back-propagates to postsynaptic sites in its dendrites. The burst learning rule identified *in vitro* (Figure 2C) is immediately invoked for all parallel fiber segments:

$$w_s \rightarrow w_s - w_s \eta_{2,4} \left[ 1 - \left( \frac{t_s - t_B}{L_{w_{2,4}}} \right)^2 \right] \quad (6)$$

where  $\eta_2$  and  $L_{w_2}$  are used if the burst of the SP cell is a small burst, and  $\eta_4$  and  $L_{w_4}$  are used if it is a large burst. Once again,  $[\dots]$  symbolize rectification, which means this rule is applied to all weights whose segment began at a time  $t_s$  as long as  $|t_s - t_B| < L_w$ . Beyond this range, the weights are unchanged.

However, burst-induced depression found *in vitro* is purely depressing and would trivially decrease all weights to zero. Therefore, a non-associative potentiating rule was added where the weights slowly relax back to  $w_{max}$  with a time constant of  $\tau_w$  according to Equation 7:

$$\tau_w \frac{dw_s}{dt} = w_{max} - w_s \quad (7)$$

This rule maintains the independence of synaptic weights and is biologically plausible since, with  $\tau_w$  sufficiently large, a weak potentiating rule would be difficult to detect experimentally. The responses of the model shown below were always quantified after weight values came to equilibrium.

### 2.5. Experimental data

Details of the surgery and recording techniques are as described previously (Marsat et al., 2009; Marsat & Maler, 2010). Briefly, craniotomy is performed under general anesthesia. During the experiment, the fish is awake but paralyzed with curare and locally anesthetized. Single-unit extracellular recordings from superficial pyramidal cells of the centro-lateral of the electrosensory lateral line lobe were performed during stimulation.

Stimuli consisted of amplitude modulations of the fish’s own electric field. The stimulus was delivered through two large global electrodes placed on each side of the fish, thereby producing a global stimulation. For local stimulation, a small dipole was placed in the center of the receptive field of the cell. The distance between the dipole and the skin was adjusted to maximally stimulate the whole receptive field of the cell while avoiding stimulation of receptors outside the classical receptive field. The intensity of both local and global stimuli were adjusted so that the modulation was 10 – 15% of the electric field of the fish as measured near the receptive field of the cell.

The difference in the SP cell response between local and global stimulation is a measure of the efficacy of the adaptive filter network. Their responses were characterized by their firing rate modulation over one AM cycle (i.e. their PSTH), their average burst rate, and a metric called cancellation. Cancellation was measured by fitting a sinusoid to the PSTH during local and global stimulation and taking the ratio of the amplitudes of the sine waves:

$$\text{Cancellation} = \left( 1 - \frac{\text{Amp}_{\text{global}}(f)}{\text{Amp}_{\text{local}}(f)} \right) \times 100\%$$

If the global modulation is identical to the local modulation, then cancellation is zero.

The experimental data with the complete model fit is given in Figures 2 and 3, as it was also shown in Bol et al. (2011). In the present paper, a primary goal, which we will introduce in the following section, is to reproduce the behavior of the model with analytical calculations, which allow for a better understanding of the dynamics occurring in the actual neural system.

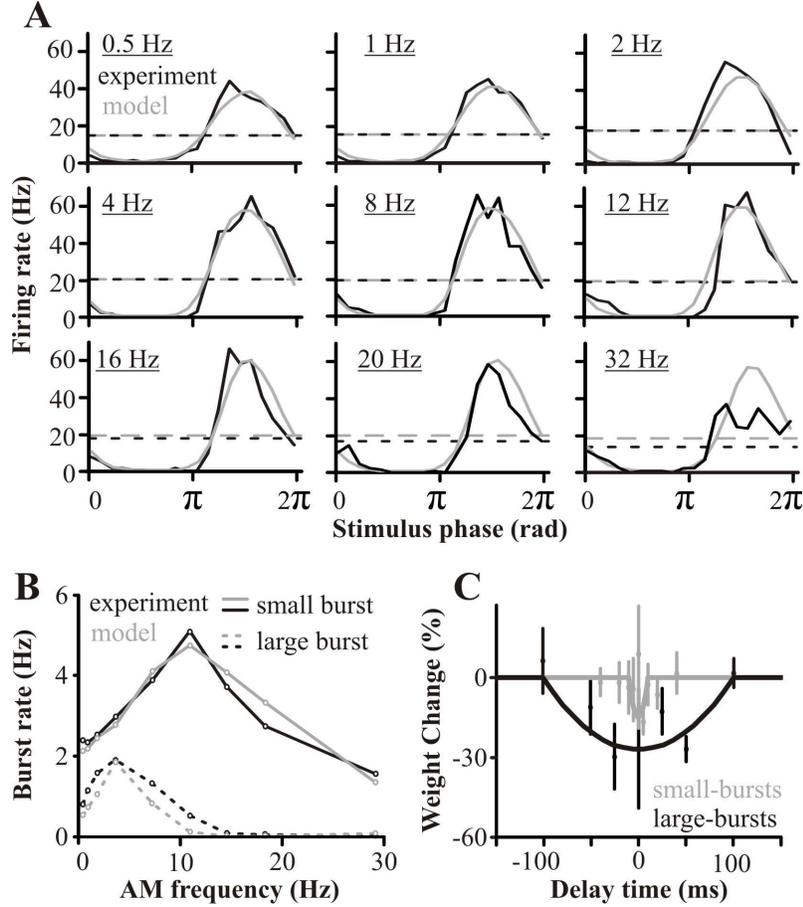


Figure 2: *In vivo* experimental data of superficial pyramidal cells when stimulated locally (black) as well as the local model fit (grey). **(A)** Peri-stimulus time histograms comparing the feedforward response of *in vivo* ( $n=9$  cells) and modeled pyramidal cells elicited by stimuli of different frequencies. Dashed lines represent the average firing rate. **(B)** Comparison of *in vivo* and modeled pyramidal cell burst rates ( $n=9$  cells) during feedforward stimulation as a function of stimulus frequency. Bursting is quantified by dividing the spike trains into small (2 or 3 spikes) or large bursts (4 or 5 spikes); longer bursts are taken as combinations of small and large bursts (see Materials and Methods). **(C)** Experimental 2-spike (grey) and 4-spike (black) burst-induced spike-time dependent synaptic plasticity at the PF-SP cell synapse as measured *In vitro* (data points). The delay time corresponds to  $t_{postsyn} - t_{presyn}$ . Also plotted is a continuous fit of the plasticity data used in the model (solid lines). Data taken from (Harvey-Girard et al., 2010).

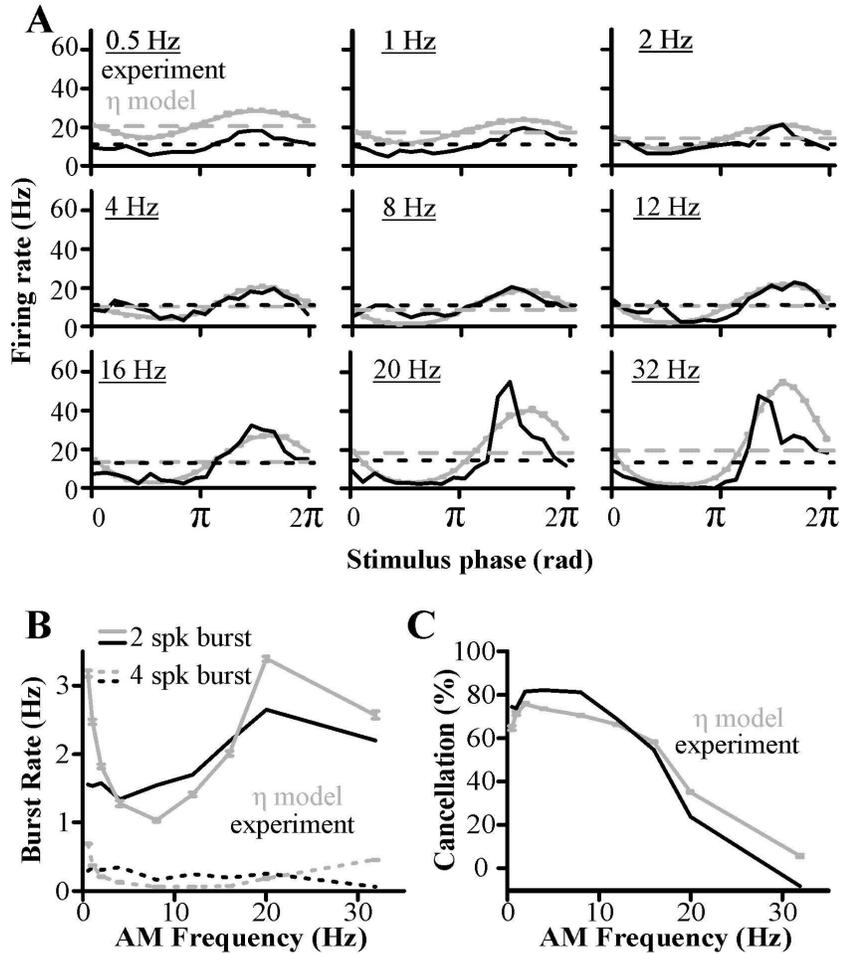


Figure 3: Comparison of *in vivo* data (black) with the model (grey) during global stimulation showing its ability to replicate the feedback induced cancellation. (A) PSTH of the model and *in vivo* responses to global stimuli of different frequencies (n=9 cells). Dashed lines represent average firing rate per second. (B) Burst rates in model and *in vivo* responses (n=9 cells). Bursting responses were segregated into small and large bursts as described previously (see also Materials and Methods). (C) Cancellation performance of the model compared to experimental data (n=9 cells; see Fig 1).

### 3. Feedforward Analytics

Generally speaking, when considering only feedforward stimulation (i.e. when the feedback pathway is not active), the firing rate of a given neuron model can be putatively calculated given the input signal and the parameter values of the model. This relationship can then be used iteratively when the feedback pathway is active to find the stable output firing rate of the system. To calculate the firing rate of a model during a slowly varying SAM input (relative to the membrane time constant of the neuron), an adiabatic approximation can be used. Thus, the input signal can be segmented into approximately independent bins with a constant input within each bin. If there are sources of positive feedback, such as from the DAP, their effect may be theoretically estimated, or alternatively, numerically solved within each bin. The specific details of each neuron model must be approached individually.

For the weakly electric fish system, the adiabatic approximation will be used to solve for the variation of the firing rate during forcing at different SAM frequencies. As with other stochastic LIF models, the system can be recast as a first passage time problem and the firing rate calculated. However, the effect of complicating factors such as low-pass filtered noise and rectification must also be addressed.

#### 3.1. ELL model

The feedforward component of the weakly electric fish model is

$$\tau_m \frac{dV}{dt} = -V + [I + \sigma \xi_L(t) + \kappa \sin(2\pi ft)], \quad (8)$$

For a constant input (i.e.  $f = 0$ ) with Gaussian white noise and without rectification, the mean firing rate,  $R$ , of this model is (Tuckwell, 1988):

$$R = \left[ \tau_{ref} + \tau_m \int_{\frac{V_r - I}{\sigma}}^{\frac{V_{th} - I}{\sigma}} \sqrt{\pi} e^{x^2} (1 + \operatorname{erf}(x)) dx \right]^{-1}, \quad (9)$$

where  $\operatorname{erf}(x)$  is the error function. To accurately model low-pass filtered noise, the mean and variance of the input in Equation 9 can be approximated by the following expressions fitted from numerical simulations:

$$I_{low} = I - \frac{0.35\sigma f_{Nyq}}{gf_{cut}} \quad (10)$$

$$\sigma_{low}^2 = \sigma^2 \frac{f_{Nyq}}{f_{cut}} \quad (11)$$

where  $f_{Nyq} = (2\tau_m)^{-1}$ . Clearly, the corrections are proportional to the ratio of the Nyquist frequency to the cutoff frequency, which is reasonable since this is the critical parameter in low-pass filtering. However, the exact form of each correction is *ad hoc*, and the only evidence that Equations 11 and 10 are appropriate is the success of the estimate. In addition, the mean input correction is to first order, and the fit deteriorates when  $\sigma > 2.5$  and  $f_{cut} < 250\text{Hz}$ . Other work on mean firing rate for filtered noise driven LIF models (e.g. Brunel et al. (2001)) can not be used here because of the rectification.

An analytical approximation to rectification can then be added by altering the mean,  $I$ , and variance,  $\sigma^2$ , of the model to account for the altered distribution of the noise. The corrections are (see Appendix: Rectification for derivation)

$$I_{rect} = \frac{I}{2} \operatorname{erfc} \left( \frac{-I}{\sqrt{2}\sigma} \right) + \frac{\sigma}{\sqrt{2\pi}} e^{-\frac{I^2}{2\sigma^2}} \quad (12)$$

$$\sigma_{rect}^2 = \frac{1}{2}(I^2 + \sigma^2) \operatorname{erfc} \left( \frac{-I}{\sqrt{2}\sigma} \right) + \frac{\sigma I}{\sqrt{2\pi}} e^{-\frac{I^2}{2\sigma^2}} - I_{rect}^2 \quad (13)$$

This provides a close fit to the model data for moderate noise intensities, even when the mean bias current is close to zero (in units of  $\sigma$ ). Deterioration does occur, however, for high noise intensities ( $\sigma > 2$ ) when the rectified noise distribution ceases to be effectively Gaussian. These values can then be substituted for  $I$  and  $\sigma$  in equations 10 and 11 to approximated both rectified input and low-pass filtered noise.

### 3.2. Positive feedback

If necessary, positive feedback - input current proportional to the output firing rate - can be included in the calculations using recursive maps. Incorporating positive feedback, the mean input to the model becomes

$$I_{eff} = I + \lambda R(I_{eff}) \quad (14)$$

where  $R(I_{eff})$  is the firing rate of the model at the new effective input current,  $I_{eff}$ , and  $\lambda$  is the strength of the feedback, which has yet to be determined. To first order,  $\lambda$  is assumed to be constant, although it too could vary with the firing rate. To find the stable firing rate, Equation 14 can be transformed into a recursive map, i.e.  $R(I_{i+1}) = R(I + \lambda R(I_i))$ , or, defining  $R_{i+1}$  as  $R(I_{i+1})$ , then  $R_{i+1} = R(I + \lambda R_i)$ . If this map is stable, then after sequential iterations the firing rate should be constant:  $R_{i+1} \approx R_i$ . However, the value of  $\lambda$  must be determined by fitting the firing rate of the model before and after positive feedback appears.

In the weakly electric fish model, the depolarizing after-potential (DAP) is a form of positive feedback. Despite the complex shape of the DAP current, a constant value of  $\lambda_{DAP}$  was identified and found to fit the model accurately.

### 3.3. Periodic input

To calculate the firing rate of the model when given an arbitrary periodic input signal (i.e.  $f \neq 0$ ), an adiabatic approximation can be used. Thus, the input can be segmented into sequential bins and the preceding methods can then be used to solve for the firing rate within each bin. This holds as long as the input is slowly varying with respect to the membrane time constant of the model cell. For the weakly electric fish model, input sinusoids of frequencies up to 32 Hz were effectively modeled with this approach (Figure 4). However, fits deteriorated for input signals greater than 40 Hz.

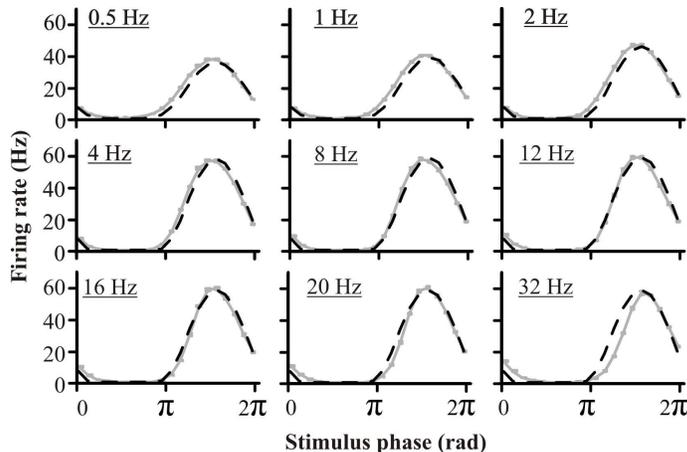


Figure 4: Comparison of the firing rate modulation of the model (grey lines) and the analytical approximation (dashed black lines) for feedforward stimulation with weak positive feedback at various input frequencies. The governing model equation is  $\tau_m \dot{V} = -V + [I + \sigma \xi_L(t) + \kappa \sin(2\pi ft)] + DAP(t)$  with parameters from Table 1.

#### 4. Feedback Analytics

In this section, the external feedback loop between synaptic weights and learning events is closed and the resulting dynamics explored. For a constant input, the equilibrium position of the synaptic weights influenced by the associative STDP rule and the non-associative potentiating rule can be calculated by determining the relationship between a) the firing rate and the learning event rates, b) the learning event rates and the stable weight value, and c) the stable weight value and the firing rate. This creates a system of equations that can be solved self-consistently. If the stimuli is slowly varying, an adiabatic approximation can again be used to segment the input into approximately constant input intervals.

##### 4.1. Learning event rate during feedback

For a given constant synaptic strength (i.e. weight value) and constant input current, the firing rate can be approximated using the same equations that were used to estimate local stimulation. If necessary, the equations can be renormalized to include changes in inhibition. Once the firing rate is known, it can be transformed into a learning event rate. For typical STDP rules, this transformation is trivial.

In the case of the weakly electric fish model, the set of possible learning events - or burst sizes - is divided for simplicity into two subsets: 2- and 4-spike bursts. Since the definition of a burst involves a threshold non-linearity, and the classification of a given spike into a 2- or 4-spike burst depends on past history, the mathematical relationship between the 2- and 4-spike burst rates and the firing rate may be difficult to calculate analytically. However, when considering only 2-spike burst events, one may employ a simplified analytical approach for the mean bursting rate given a fixed input.

Attending to Equation 5, one can observe that the variable  $b$  experiences an instantaneous increase every time a spike occurs, and after that, an exponential

decay follows. Considering low firing rates, we can assume that, prior to the arrival of a given spike, the variable  $b$  is in its resting state ( $b = 0$ ). The incoming spike induces an increase of  $A$  in  $b$ , after which  $b$  starts to decay. If a 2-spike burst occurs, then from our definition of a 2-spike burst a second spike should arrive within the next 15  $ms$  after this first spike (see Figure 5A). The second spike will increase  $b$  to a higher value than the first one, due to the residual value of  $b$  from the first spike and to the nonlinearity in Equation 5. It is easy to see that the minimal value reached by  $b$  after the second spike, namely  $b_{th}$ , corresponds to the situation in which such a spike arrives exactly 15  $ms$  after the first one (black line in Figure 5A), and we can obtain

$$b_{th} = A(1 + \exp(-h/\tau) + A B \exp(-2h/\tau)), \quad (15)$$

where  $h = 15 \text{ ms}$  is the temporal window for the burst. If the arrival of the second spike is less than 15  $ms$  after the first spike, the maximum in  $b$  will be higher, clearly surpassing  $b_{th}$  (gray line in Figure 5). Therefore, the value  $b_{th}$  constitutes a threshold for the occurrence of 2-spike bursts in our model. In order to avoid a miscounting of the number of 2-spike bursts we assume that, for theoretical purposes only,  $b$  resets to zero after the occurrence of a burst. This is approximately valid when bursts can be considered as rare events (i.e. when dealing with relatively low firing and bursting rates), and it is the case in our system when the feedback pathway is active.

From the values of the parameters that we adopted, we have that  $b_{th} \simeq 0.67$ . Since  $b$  is restricted to have values lower than  $b_{th}$  because of our reset condition (and in practice it will be much lower most of the time), one may neglect the nonlinear term  $Bb^2 \sum_n \delta(t - t_n)$  in Equation 5 for simplicity. In addition, by averaging over the population of SP cells, we can replace the sum over spikes by the instantaneous mean firing rate  $\nu(t)$ . Finally, since we are considering a constant input, we may approximate the instantaneous mean firing rate as  $\nu(t) \simeq \nu_0 + \sqrt{\tau}\sigma_\nu\eta(t)$ , where  $\nu_0$  is the time-averaged mean firing rate,  $\sigma_\nu$  is its corresponding standard deviation, and  $\eta(t)$  is a Gaussian white noise. The resulting equation, given by

$$\tau \frac{db}{dt} = -b + \tau A \nu_0 + \tau^{3/2} A \sigma_\nu \eta(t), \quad (16)$$

together with the threshold and reset conditions, can be treated following standard procedures (Tuckwell, 1988), and therefore the mean 2-spike bursting rate for the neuron model is obtained:

$$Br_2 = \left[ \tau \int_{-\nu_0/\sigma_\nu}^{\frac{b_{th} - A\tau\nu_0}{A\tau\sigma_\nu}} \sqrt{\pi} \exp(x^2)(1 + \operatorname{erf}(x)) dx \right]^{-1}. \quad (17)$$

In order to complete the description, one has to give values for  $\nu_0$  and  $\sigma_\nu$ . For the time-averaged mean firing rate, two approaches are possible: 1) to consider a recursive map as we showed earlier, or 2) to simplify the DAP mechanism so that one may obtain an approximated theoretical expression. This second possibility can be now considered due to the presence of feedback, which induces lower firing rates on the system.

An interesting simplification, but not the only one, is to assume that the feedback due to the DAP occurs in a very short amount of time. This may be

taken into account by using the expression for  $DAP(t)$  proposed in Laing & Longtin (2002), that is,

$$DAP(t) = \alpha_L b(t) \sum_n \Theta(t_n - t_{n-1} - r_L) \delta(t - t_n - \sigma_L), \quad (18)$$

where  $\Theta(x)$  is the Heaviside step function. The above expression assumes that the positive feedback from DAP mechanisms has an instantaneous strength  $\alpha_L b(t)$  which arrives at the soma of the neuron after a delay  $\sigma_L$  from the last spike time, providing that the last inter-spike interval was larger than  $r_L$ . We can employ Equation 18 instead of Equation 2 in order to obtain a theoretical approach to DAP. To ensure that the firing and bursting rates are the same as before, the new parameters are fitted to  $\alpha_L = 0.8$ ,  $\sigma_L = 9 \text{ ms}$  and  $r_L = 0.7 \text{ ms}$ .

We will consider here that inter-spike intervals are larger than  $r_L$  (which is true for moderate input intensities such as the ones considered in this work). In such conditions, it is easy to see that the effect of the DAP, according to this model, is the instantaneous increase of  $\alpha_L b(t)$  in the membrane potential at time  $t$ . Since we are interested in the DAP caused by the first spike in a given burst (since it will be relevant for the burst occurrence), and since we have  $b(0) = A$  with the first spike occurring at  $t = 0$ , the increase in  $V(t)$  due to DAP will arrive at  $t = \sigma_L$  to the soma, and its strength will be  $\alpha_L A \exp(-\sigma_L/\tau)$ . Finally, assuming that the delay  $\sigma_L$  is comparable to the neural refractory period, one may simply identify the increment due to DAP with an extra term in the reset potential. Therefore, the time-averaged mean firing rate can be obtained from Equation 9 by considering  $V_r = \alpha_L A \exp(-\sigma_L/\tau)$  instead of  $V_r = 0$ .

The standard deviation  $\sigma_\nu$ , on the other hand, is more difficult to treat. This is due mainly to the fact that the bursting dynamics induces nontrivial temporal correlations in the spike times, and therefore our previous approximation  $\nu(t) \simeq \nu_0 + \sqrt{\tau} \sigma_\nu \eta(t)$  with  $\eta(t)$  being a Gaussian white noise is not precise enough. Similarly as in the previous case of low-filtered noise, we employed here a numerical fit which takes into account the effects of temporal correlations on the noise strength, so that we can still preserve the white noise approximation by rescaling  $\sigma_\nu$ . In this case, the noise strength also presents a dependence on the input  $I$ , since a higher input induces more bursts (see caption of Figure 5 for details).

The comparison between these theoretical estimations and the numerical results of the model is shown in Figure 5B. To obtain both the theoretical and numerical data, mean firing rate and mean 2-spike burst rate were measured for different values of the input bias  $I$ , which was varied in a range of  $0.5 \sim 1.5$ . As the Figure 5B shows, our theoretical estimation for the mean 2-spike bursting rate is in agreement with simulations of the model. In addition, the inset shows the good agreement of our estimation for  $\nu_0$  (taking into account the correction of  $V_r$  due to the DAP) with the numerical simulations.

Unfortunately, the theory can not be extended to the case of 2- and 4-spike bursts, or even 4-spike bursts only, because of extra difficulties in defining a threshold for large bursts and discriminating whether a given spike belongs to a small or a large burst. For these situations, we are restricted then to fit the relationship between firing rate and bursting rate from numerical simulations of the model. However, the theory can be used effectively when only 2-spike bursts are considered, and the results found in the following sections are the same when

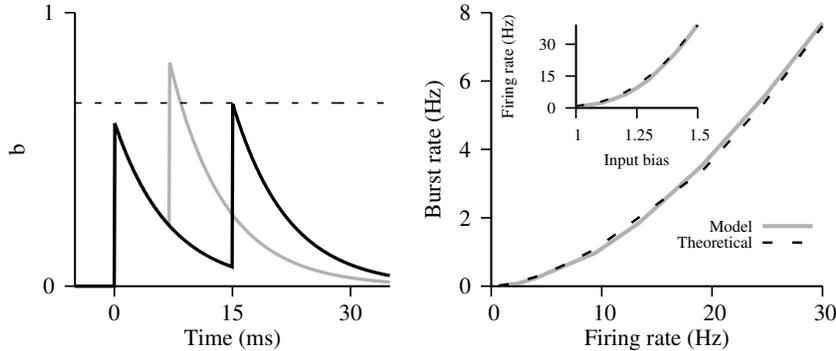


Figure 5: Theoretical approach of 2-spike bursting rate. (A) Scheme of the dynamics of  $b$  and the effects of spikes on it. The threshold for 2-spike bursts is depicted as a dashed line, whereas black and grey lines corresponds to different example trials. (B) Relationship between mean firing rate and mean 2-spike bursting rate, according to numerical simulations of the model (grey line) and our theoretical approach (dashed line). The standard deviation of the mean firing rate was fitted to  $\sigma_\nu = 0.45I - 0.095$ . Inset: Mean firing rate as a function of the input bias  $I$ . The theoretical approach (which includes the effect of DAP) is compared with simulations (grey line).

using this theory and when using a fit from numerical simulations. When both small and large bursts are considered, identifying 2- and 4-spike bursts leads to a non-monotonic 2-spike burst relationship because of the mutual exclusivity of spikes among bursts (and 4-spike bursts take precedence).

#### 4.2. Learning event impact

In this section, the effect of each learning event on the average weight value is calculated. During a constant input, the feedback pathway may still be periodic with frequency  $f$  because of intrinsic periodicity of the feedback neurons. Alternatively, there may be only a single effective synapse during constant input, but the calculations performed here will still hold. Although the timing of each pyramidal cell learning event is stochastic, the effect of each event on the average weight value can be calculated because the firing activity of the feedback is predictable. From the definition of the mean, the difference in the average weight value,  $\bar{w}$ , before and after a learning event is equal to

$$\bar{w}^+ - \bar{w}^- = \frac{\sum_s w_s^+}{n_w(f)} - \frac{\sum_s w_s^-}{n_w(f)}$$

where  $w_s^-$  and  $w_s^+$  denote the weight values just before and after a learning event, respectively, and  $n_w(f)$  is the number of active and modifiable feedback paths that each repeat at frequency  $f$ . If learning is governed by an additive STDP rule then

$$w_s^+ - w_s^- = STDP(t_E - t_s),$$

where  $t_E$  is the time of the learning event,  $t_s$  is the start time of segment  $s$  and  $STDP(t_E - t_s)$  is the learning rule as a function of the pairing delay between these occurrences. Substituting the learning rule for  $w_s^+$  in the previous

equation, this becomes

$$\begin{aligned}\bar{w}^+ - \bar{w}^- &= \frac{\sum_s \{w_s^- + STDP(t_E - t_s)\}}{n_w(f)} - \frac{\sum_s w_s^-}{n_w(f)} \\ &= \frac{\sum_s STDP(t_E - t_s) T/n_w(f)}{T}\end{aligned}$$

Since the average duration of each feedback delay line is  $T/n_w(f)$ , the numerator can be thought of a Riemann sum estimation of the area of the STDP rule. If the plasticity rule is additive, this estimation is accurate. This approximation can still hold even if the plasticity rule is multiplicative as long as the spread of the weight distribution is small at equilibrium (i.e. when all weight segments are identical and when the impact of each learning event is small compared to the average weight value). Therefore,  $\bar{w}$  can be substituted for the weight at each segment and factored out of the numerator. Note that for a given STDP rule, the relative impact of one burst on the average weight value increases as the AM frequency increases.

For the weakly electric fish model, the associative rule is multiplicative and quadratic with the pairing delay, so the learning event impact becomes:

$$\bar{w}^+ - \bar{w}^- \approx \frac{-\bar{w}^- \eta \int_{-L_w}^{L_w} \left(\frac{t^2}{L_w^2} - 1\right) dt}{T} \quad (19)$$

$$\Delta \bar{w} \approx \frac{-4\bar{w}\eta L_w}{3T} \quad (20)$$

Here,  $\eta$  and  $L_w$  would be substituted for the appropriate 2- or 4-spike burst values to determine the effect of each burst.

#### 4.3. Average weight and learning event rates

When the model is in equilibrium, the weights fluctuate around their values as occasional learning events depress them while the homeostatic potentiation rule slowly increases them. Schematically, the weight dynamics might look like Figure 6.

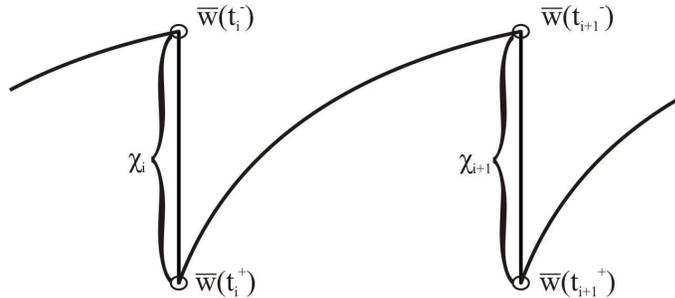


Figure 6: Schematic of the dynamics of the average weight value at equilibrium. Learning events occur at time  $t_i$  and  $t_{i+1}$ , which depress the value of the average weight by an amount  $\chi_i$  and  $\chi_{i+1}$ , respectively. The weight values then slowly potentiate according to  $\tau_w \dot{\bar{w}} = w_{max} - \bar{w}$ . A difference map can be defined that transforms the average weight just after learning event  $i$  to the average weight just after event  $i + 1$ :  $\bar{w}(t_i^+) \rightarrow \bar{w}(t_{i+1}^+)$ .

Thus, the average weight value at time  $t$  is defined as  $\bar{w}(t)$  and the depression of  $\bar{w}(t)$  after learning event  $i$  is equal to  $\chi_i$ . This system can be analyzed as

an iterative mapping of weight values just after a learning event occurs. If a learning event arrives at time  $t_i$ , then

$$\bar{w}(t_i^+) = \bar{w}(t_i^-) - \chi_i \quad (21)$$

where  $t_i^-$  is defined as the time just before the event,  $t_i^+$  is defined as the time just after the event, and  $\chi_i$  is the effect of one learning event on the average weight value at time  $t_i$ . Between events, the weights are only under the influence of the potentiation rule. If the difference between weights at different segments is small,  $\bar{w}(t_{i+1}^-)$  can be found by solving the homeostatic rule:

$$\begin{aligned} \tau_w \frac{d\bar{w}}{dt} &= w_{max} - \bar{w} \\ \Rightarrow \bar{w}(t_{i+1}^-) &= w_{max} + (\bar{w}(t_i^+) - w_{max}) e^{\frac{-(t_{i+1}^- - t_i^+)}{\tau_w}} \end{aligned}$$

Defining the time difference  $t_{i+1}^- - t_i^+$  as  $\Delta t$  and substituting this into Equation 21 gives

$$\bar{w}(t_{i+1}^+) = w_{max} + (\bar{w}(t_i^+) - w_{max}) e^{\frac{-\Delta t}{\tau_w}} - \chi_i$$

If this system is in equilibrium, then  $\bar{w}(t_{i+1}^+) \approx \bar{w}(t_i^+) \approx \bar{w}$ , which is the equilibrium average weight value within a bin. This is a good approximation as long as  $\chi_i/w \ll 1$ . The time difference  $\Delta t$  is the time between successive bursts during equilibrium, so the reciprocal of  $\Delta t$  is just the equilibrium learning event rate in that bin. During the actual simulation, the events will not arrive exactly  $\Delta t$  apart due to noise but if the weights stay near their equilibrium value, then  $\Delta t$  should be the mean time between learning events.

In addition,  $\chi_i$  is the effect on the average weight value of one event, which can be calculated using the approach in the previous section. Furthermore, at equilibrium the impact of learning events at different times is identical ( $\chi_i = \chi_{i+1} = \chi$ ). Defining  $A$  as the area of the STDP rule, then  $\chi = \bar{w}A$  for a multiplicative rule, and solving for  $\bar{w}$  gives

$$\bar{w} = \frac{w_{max} \left(1 - e^{\frac{-1}{\tau_w E}}\right)}{1 - e^{\frac{-1}{\tau_w E}} + A} \quad (22)$$

This equation relates the equilibrium event rate,  $E$ , to the equilibrium weight value,  $\bar{w}$  during a constant input. For the electric fish model with 2-spike bursts only, this equations becomes

$$\bar{w} = \frac{w_{max} \left(1 - e^{\frac{-1}{\tau_w B r_2}}\right)}{1 - e^{\frac{-1}{\tau_w B r_2}} + \frac{4\eta_2 L w_2}{3T}}, \quad (23)$$

#### 4.4. Multiple Learning Events

The preceding analysis can be easily extended to multiple learning events if the different events have sufficient separation in impact and probability. In this case, the most powerful, rarest (i.e. major) events can be assumed to maintain the weights around equilibrium as in the previous section, and the weaker, more common (i.e. minor) events can be included as a correction to the potentiating

rule between major events. Note that if major events were more probable *and* more powerful, then this technique will still work (although in this case, the minor events could likely be ignored). If the potentiating rule is weak, then the effect of the individual timing of each minor event between sequential major events is negligible. Therefore, the total effect of  $n$  minor events between 2 major events can be added at time  $t_{i+1}^-$ , when the next major event occurs. Furthermore, the average value of  $n$  can be calculated as the ratio of the minor event rate,  $E_{minor}$ , over the major event rate  $E_{major}$ . Solving again for  $\bar{w}$  gives

$$\bar{w} = \frac{w_{max} \left( 1 - e^{\frac{-1}{\tau_w E_{major}}} \right)}{1 - e^{\frac{-1}{\tau_w E_{major}}} + A_{major} + \frac{E_{minor}}{E_{major}} A_{minor}} \quad (24)$$

where  $A_{major}$  and  $A_{minor}$  are the areas of the major and minor learning event STDP rules (assuming both are multiplicative rules), respectively. For the electric fish model, 4-spike bursts are major learning events (with event rate  $Br_4$ ) while 2-spike bursts are minor events (with event rate  $Br_2$ ), and this equation becomes

$$\bar{w} = \frac{w_{max} \left( 1 - e^{\frac{-1}{\tau_w Br_4}} \right)}{1 - e^{\frac{-1}{\tau_w Br_4}} + \frac{4\eta_4 L_{w_4}}{3T} + \frac{4Br_2\eta_2 L_{w_2}}{3Br_4T}}$$

#### 4.5. Periodic input

In a similar approach to the local theoretical analysis of the PSTH, a slowly varying input can be divided into sequential bins and the preceding equations can be solved for the specific input current within each bin.

Summing all these individual relationships together, we have a set of equations that must be self consistent: the weight value in each bin must create a firing rate (via Equation 9) that produces a learning event rate (via Equation 17 or numerical fits), that generates a weight value (via Equation 23 or 24) that, in equilibrium, is identical to the initial weight value input. This can be solved using a root finding algorithm and must be iterated for each bin.

For the weakly electric fish model, the governing equation within each bin is

$$\tau_m \frac{dV}{dt} = -V + [I + \sigma \xi_L(t) + \kappa \sin(2\pi f t^\phi)] + DAP(t) + \Lambda (w_s^\phi - gV), \quad (25)$$

where  $t^\phi$  is the time in the middle of that particular bin. Within each bin, there are  $n_w(f)$  feedback segments that each repeat with a given frequency  $f$  identical to the input frequency. On the other hand, since the input within each bin is not modulated, all of the weights,  $w_s^\phi$ , will converge to the same average value and only  $\bar{w}^\phi$  is important.

For the weakly electric fish model with 2-spike bursts only, this method is accurate in predicting the behavior of the model (Figure 7). However, when including 2- and 4-spike bursts, the theoretical approximation deteriorates at high AM frequencies ((Figure 8). This is due to the breaking of the adiabatic approximation for 4-spike bursts. 4-spike bursts can, by definition, form within 45 ms, and this is on the order of a period of a high frequency input. Therefore, 4-spike bursts are certainly not isolated to individual bins, and the quasi-equilibrium approach fails.

This final model also includes a decrease in the learning rate  $\eta$  at high frequencies to fit experimental data. A decreased learning rate theoretically necessitates more bursting, but this fails to materialize in the model due to the finite time required to burst. If  $\eta$  is kept constant, the fit between model and analytics is much better (Figure 8C).

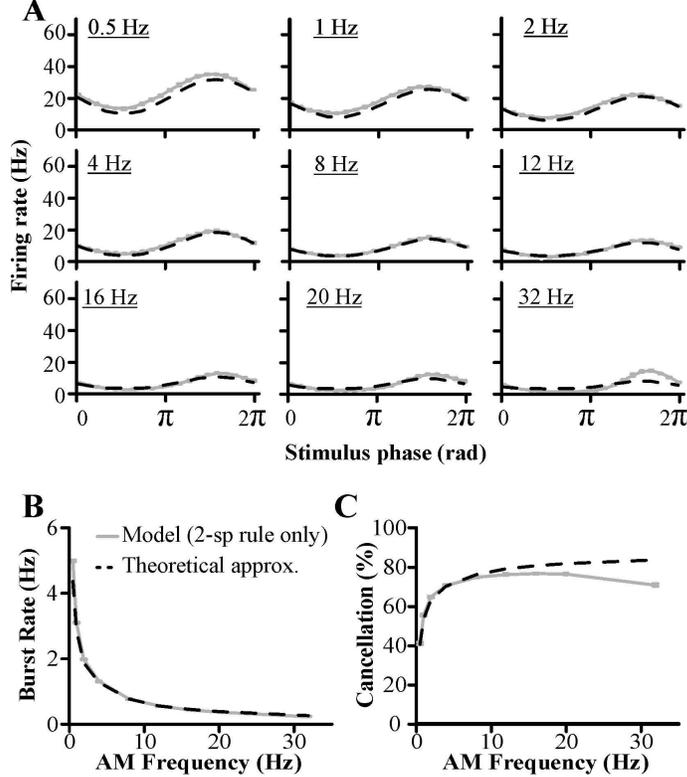


Figure 7: Comparison of the (A) PSTH, (B) burst rate, and (C) cancellation of the model (grey solid lines) and the analytical approximation (dashed black lines) during global stimulation with only the 2-spike burst rule active. The analytical approximation was generated by solving Equations 9 and 22 self-consistently within individual segments of the AM period and transforming the firing rate to a 2-spike burst rate using a numerically derived equation. Here,  $g$  was changed to 1.66 so that the average firing rate of the model and experiment at 4 Hz are similar. In this simulation  $\eta$  is constant across all frequencies.

## 5. Discussion

We have provided a detailed analysis of a biophysically plausible model of redundant signal cancellation. At its core is a burst-time dependent STDP rule, and in the particular application of interest, the weakly electric fish *Apteronotus leptorhynchus*, this rule is a correlative depressing rule. The importance of bursts as learning events makes their simulation important in such a network, which goes beyond previous rate-based modeling attempts (Bastian et al., 2004). The analytical estimation of burst event production in a simple LIF type model poses serious challenges, and our approximations are accurate and can be usefully applied in other contexts. In particular, we have provided a way to estimate

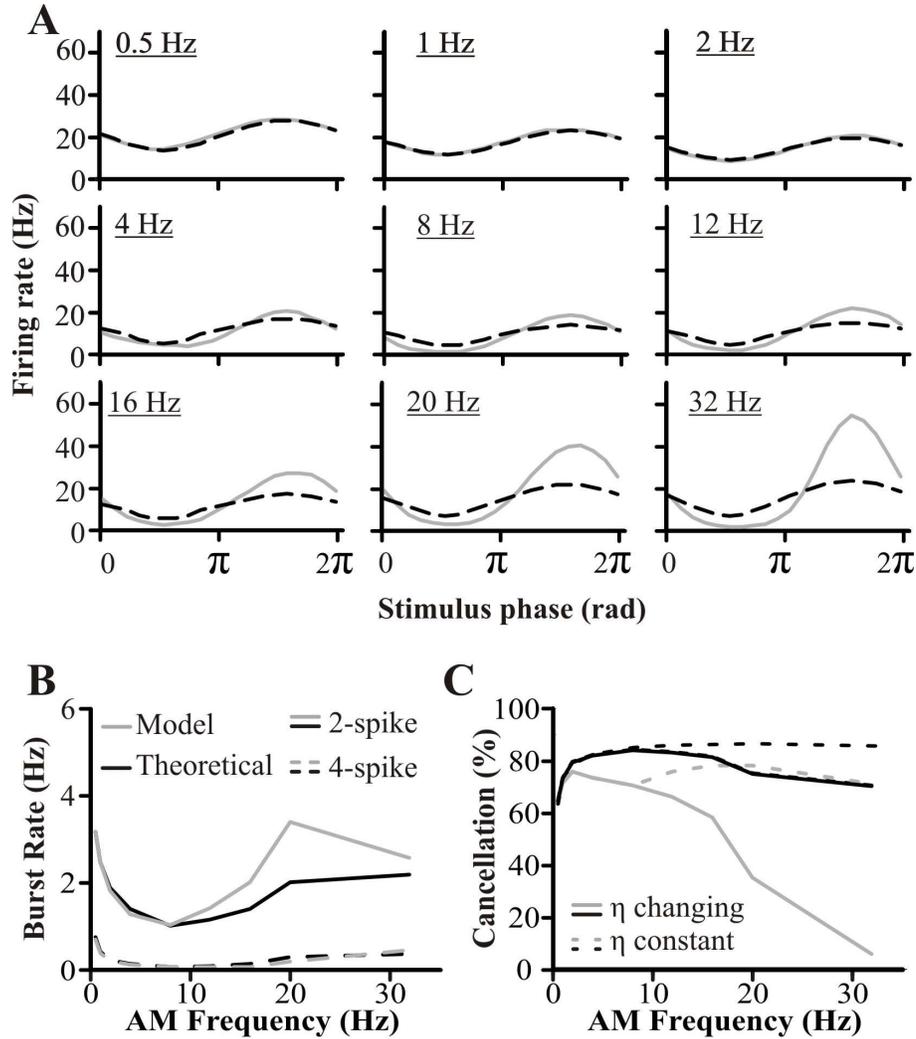


Figure 8: Comparison of the (A) PSTH, (B) burst rate, and (C) cancellation of the model (gray lines) and the analytical approximation (black lines) during global stimulation with both 2- and 4-spike burst rules active. In this simulation,  $\eta$  was altered at high AM frequencies to fit to experimental data. Also plotted in C is the cancellation of the model and the analytical fit when  $\eta$  is constant across all frequencies (dotted lines). The analytical approximation was generated by solving Equations 9 and 22 self-consistently within individual segments of the AM period and transforming the firing rate to a 2- and 4-spike burst rate using numerically derived equations.

these rates in a feedback context using a self-consistent firing and burst rate formalism. This enabled us to derive the behavior of the learning weights as well. The analytical treatment of our model, which has been developed here for the first time, yields results which agree generally very well with those from direct numerical simulations of the model. In turn, those simulations were shown recently to reproduce well the experimental observations (Bol et al., 2011), thus our theoretical framework enables a deeper understanding of the mechanisms that generate redundancy cancellation.

During local stimulation, the firing rate of the stochastic LIF model was found by solving a well-known first passage time problem (Equation 9) (Doiron et al., 2004). However, the effect of filtering the noise and rectifying the input to a stochastic LIF model has not been thoroughly investigated and the corrections to Equation 9 outlined in this paper are novel. Although the rectified input correction was a simple approximation of a Gaussian distribution to a non-Gaussian density, the theoretical foundations of the low-pass filtered noise corrections are more ambiguous. These corrections are based on the ratio of the cutoff frequency to the Nyquist frequency of the numerical simulation, which is the critical parameter in low-pass filtering. It is understandable that the variance of the noise in the FPT equation might be augmented to accommodate low-pass filtered noise, but it is harder to explain why a mean correction is necessary or why it has the form it does.

Negative feedback has already been investigated analytically in stochastic LIF models and solved by calculating the effective input current,  $I_{eff}$ , and making the process stationary to find the equilibrium firing rate (Sutherland et al., 2009). A similar approach was used to study the dendritic after-potential (DAP), a known source of positive feedback for superficial pyramidal neurons in the ELL. Despite its complexity, reducing the DAP to a single feedback parameter on the firing rate was a successful approximation in this model given the parameters that mimic experimental local stimulation. On the other hand, we also employed an alternative approach, which consisted on treating the positive DAP feedback as an instantaneous impulse arriving to the soma of the neuron shortly after a spike occurred. This approach also gave good results when low firing rates were considered, as it is the case when feedback pathway is active. Both the recursive approach and the impulse approach are novel methods to analyze the dendritic after-potential introduced in this work.

The biological system that the ELL model simulates is unique in that the learning events are a function of the bursting rates, and not simply of the firing rate itself. This complicates the analysis considerably, and the best strategy that we could follow when different learning events were considered is to employ purely numerical analysis. We have found, however, a novel solution for the case when only short learning events are considered. The theoretical solution developed here is general enough to be used in a wide variety of neural systems, and in the case of the electric fish it allows to theoretically estimate the learning event rates for 2-spike bursts. Unfortunately, this approach breaks down when other learning events are considered, although one may employ the intuition provided by our theoretical approach to better understand the dependence of bursts with firing rate in more complex situations.

It is interesting to mention that the concrete definition of burst is not of vital importance in our model. The one chosen here, *i.e.* the occurrence of a number of spikes within a fixed time window, is analytically and computationally

adequate, but it is also consistent from a biophysical point of view. This is due to the fact that (i) the ISI distribution of the SP cells is bimodal (Bol et al., 2011), with a frequency-independent peak at very small values of ISI indicating the presence of bursting (Longtin & Chialvo, 1998), and (ii) single-pulse paired stimulation does not trigger plasticity in our system (Harvey-Girard et al., 2010). These two conditions together imply that bursts in SP cells appear as well time-located events and functionally different from single spikes, and these conditions are embodied in our model and are important for its analysis.

There is ongoing debate concerning the propensity of the granule cells to fire in bursts (Chadderton et al., 2004; Rancz et al., 2007; D’Angelo et al., 1998, 2001; Jorntell & Ekerot, 2006). In the present study, we are dealing with a specialized group of cerebellar cells, the Zebrin-2 negative cells (Brochu et al., 1990), whose firing propensity has not been studied to date. However, the propensity for bursting of granule cells is not a strong requirement for our conclusions to hold. Indeed, the periodic nature of the global sensory input is likely to induce the clustering of granule cell spikes around a range of phases of the input; this is so even if these cells do not have a propensity for bursting via an intrinsic burst mechanism. In this sense, the only requirement is that the range of firing phases of the granule cells scales with the period of the stimulus (just as it occurs for superficial neurons), and such a feature is plausible for both a bursting-propensity or a single spike-propensity scenario.

Adaptive filtering is further complicated by the interaction between the learning event impact and signal frequency. Each learning event (be it single spike or burst) has a well-defined effect on the synaptic strengths of the feedback regardless of the stimulus frequency. Importantly, the temporal width of the learning rule (i.e.  $L_w$ , the maximum pre-post delay to still cause plasticity) is fixed. On the other hand, the feedback must be phase-locked to the stimulus for cancellation to occur, so the proportion of weights affected by a single learning event (i.e.  $L_w/T$ ) will change with the stimulus period. For the behavior of the model to remain constant (i.e. unaltered firing and burst rates), the learning event rate must linearly decrease with stimulus frequency. The generation of learning events is not strongly affected by the stimulus frequency, however, which causes  $\bar{w}$  and the learning event rates to change at different frequencies. As the average firing rate of the neuron is directly related to the learning event rate and to the input sensitivity of the neuron, altering the firing rate at different stimulus frequencies is an unwanted effect.

In addition, a decreased event impact causes cancellation to deteriorate at low frequencies. This is due to the non-linear relationship between input current and firing rate. As the AM frequency decreases, the weights will overall potentiate because of the reduced impact of each learning event, as explained above. On the other hand, the variation of the weight distribution is not strongly affected. This leads to an approximately equal increase in current at all phases of the input stimulus. However, this equal input increase will not create equal firing rate increases at all phases of the AM stimulus. In the trough of the AM, where the input is initially low, the firing rate curve has a shallow slope, and an increase in current will not strongly change the firing rate. In the crest of the AM, the input is already high and a further increase in input will cause a large change in firing rate. Since the firing rate of the crest increases more than the trough, the amplitude of the PSTH increases and the cancellation decreases.

This feature might be the reason why burst-induced plasticity occurs at

the PF-SP cell synapse in *A. leptorhynchus* and not plasticity based on single spikes. With a learning event (burst or single spike or otherwise) having different effects at different frequencies due to the periodic nature of the feedback, a single learning rule is insufficient to maintain a constant level of synaptic depression. A solution is to recruit other learning events that become more probable as the learning event impact wanes. Higher order burst rules (3-3, 4-4, 5-4, etc) are ideal for this role since each rule will uniquely affect different frequencies and are only likely at low frequencies when the strength of each learning event is small. A synapse based on single spike plasticity does not have the flexibility to add these rare but powerful learning events.

In addition, pyramidal cell bursts have been shown to selectively encode low frequency stimuli (0-16 Hz), such as prey, to higher brain centers (Oswald et al., 2004). Burst-driven cancellation is optimal below 16 Hz (Bol et al., 2011) and would minimize bursts in superficial cells induced from predictable stimuli, reducing the noise in the putative burst channel. This is an optimal configuration to detect low-frequency stimuli even during a global predictable stimulus and provides further support to the notion that bursts are separate important information messengers.

Finally, the theory presented here enables the analysis of many combined effects on spiking dynamics, namely, periodic input, intrinsic bursting, intrinsic noise, and delayed spiking feedback along with a modulation of feedback strength by (burst-)STDP. Together with frequency-dependent cerebellar channels, this complex arrangement appears necessary to perform redundancy reduction online, i.e. without knowing what the stimulus is. This is in contrast to the relatively simpler circuitry (based on anti-Hebbian STDP) for the cancellation of inputs when the cells involved receive corollary discharge (Roberts & Bell, 2000; Roberts & Portfors, 2008). In this latter case, the mechanism could work even in the absence of redundancy, although in practice the pulses are quite regularly emitted. It remains to be seen how the mechanisms elucidated for our case could be adapted to cancel signals with little or no redundancy.

## Appendix A. Rectification

Rectification of the input complicates the analytical approximation since the input distribution ceases to be Gaussian around the mean input,  $I$ , but becomes a mixed distribution that has a continuous Gaussian probability density for values greater than 0, and a discrete probability mass at 0. In other words, if  $Y$  is the feedforward input to the model, then the probability density of  $Y$  is

$$f(y) = \begin{cases} 0 & y < 0 \\ \int_{-\infty}^0 \frac{1}{\sqrt{2\pi}\sigma} e^{-(y-I)^2/2\sigma^2} dy & y = 0 \\ \frac{1}{\sqrt{2\pi}\sigma} e^{-(y-I)^2/2\sigma^2} & y > 0 \end{cases}$$

In general, the first passage time derivation of the firing rate that produced Equation 9 will not hold when the noise is rectified. However, if the probability mass at  $y = 0$ , is small enough, then Equation 9 should still approximate the firing rate. Certainly, when the mean is large compared to the variance, then the distribution of  $Y$  will be far from 0 and  $Y$  will retain most of its Gaussianity. For intermediate values of the mean and variance when there is a non-negligible probability mass at  $y = 0$ , then Equation 9 could be used if the distribution of

the input is still assumed to be Gaussian but with the mean and variance of  $Y$  as parameters instead of  $I$  and  $\sigma^2$ . The mean value of  $Y$  is

$$E\{Y\} = 0 + \int_0^\infty y \frac{1}{\sqrt{2\pi}\sigma} e^{-(y-I)^2/2\sigma^2} dy$$

$$E\{Y\} = \frac{I}{2} \operatorname{erfc}\left(\frac{-I}{\sqrt{2}\sigma}\right) + \frac{\sigma}{\sqrt{2\pi}} e^{-\frac{I^2}{2\sigma^2}}$$

where  $E\{\}$  is the expectation operator. Similarly, the expected value of  $Y^2$  is

$$E\{Y^2\} = 0 + \int_0^\infty y^2 \frac{1}{\sqrt{2\pi}\sigma} e^{-(y-I)^2/2\sigma^2} dy$$

$$E\{Y^2\} = \frac{1}{2}(I^2 + \sigma^2) \operatorname{erfc}\left(\frac{-I}{\sqrt{2}\sigma}\right) + \frac{\sigma I}{\sqrt{2\pi}} e^{-\frac{I^2}{2\sigma^2}}$$

and so  $\sigma_y^2 = E\{Y^2\} - E\{Y\}^2$  can be calculated. Using the mean and standard deviation of  $Y$  as  $I$  and  $\sigma$  and then applying the low-pass filtered noise corrections produces the analytical estimate of rectified input to an LIF model with low-pass filtered noise:

$$I_{rect} = E\{Y\} - \frac{0.35\sigma_y f_{Nyq}}{g f_{cut}} \quad (\text{A.1})$$

$$\sigma_{rect}^2 = \frac{\sigma_y^2}{2\tau_m f_{cut}} \quad (\text{A.2})$$

This provides a close fit to the model data for moderate noise intensities, even when the mean bias current is close to zero (in units of  $\sigma$ ). Deterioration occurs for high noise intensities ( $\sigma > 2$ ) when the rectified noise distribution ceases to be effectively Gaussian.

## References

- Babineau, D., Longtin, A., & Lewis, J. E. (2006). Modeling the electric field of weakly electric fish. *J Exp Biol*, *209*, 3636–3651.
- Barlow, H. (2001). Redundancy reduction revisited. *Network: Comp Neur Sys*, *12*, 241–253.
- Bastian, J., Chacron, M., & Maler, L. (2004). Plastic and nonplastic pyramidal cells perform unique roles in a network capable of adaptive redundancy reduction. *Neuron*, *41*, 767–779.
- Bell, C., Han, V., Sugawara, Y., & Grant, K. (2000). Synaptic plasticity in a cerebellum-like structure depends on temporal order. *Nature*, *387*, 278–281.
- Berman, N., & Maler, L. (1999). Neural architecture of the electrosensory lateral line lobe: adaptations for coincidence detection, a sensory searchlight and frequency-dependent adaptive filtering. *J Exp Biol*, *202*, 1243–1253.
- Bol, K., Marsat, G., Harvey-Girard, E., Longtin, A., & Maler, L. (2011). Frequency-tuned cerebellar channels and burst-induced LTD lead to the cancellation of redundant sensory inputs. *J. Neurosci.*, *31*, 11028–11038.

- Brochu, G., Maler, L., & Hawkes, R. (1990). Zebrin-II - a polypeptide antigen expressed selectively by Purkinje-cells reveals compartments in rat and fish cerebellum. *J. Compar. Neurol.*, *291*, 538–552.
- Brunel, N., Chance, F., Fourcaud, N., & Abbott, L. (2001). Effects of synaptic noise and filtering on the frequency response of spiking neurons. *Phys. Rev. Lett.*, *86*, 2186–2189.
- Carr, C. E., & Maler, L. (1986). Electroreception in gymnotiform fish: central anatomy and physiology. In T. H. Bullock, & W. Heiligenberg (Eds.), *Electroreception* (pp. 319–374). New York: Wiley.
- Chacron, M., Longtin, A., & Maler, L. (2011). Efficient computation via sparse coding in electrosensory neural networks. *Curr. Opin. Neurobiol.*, *21*, 752–760.
- Chacron, M., Longtin, A., St-Hilaire, M., & Maler, L. (2000). Suprathreshold stochastic firing dynamics with memory in P-type electroreceptors. *Phys. Rev. Lett.*, *85*, 1576–1579.
- Chadderton, P., Margrie, T., & Hausser, M. (2004). Inhibitory feedback required for network burst responses to communication but not to prey stimuli. *Nature*, *428*, 856–860.
- Chen, L., House, J. L., Krahe, R., & Nelson, M. E. (2005). Modeling signal and background components of electrosensory scenes. *J Comp Physiol A*, *191*, 331–345.
- D’Angelo, E., De Filippi, G., Rossi, P., & Taglietti, V. (1998). Ionic mechanism of electroresponsiveness in cerebellar granule cells implicates the action of a persistent sodium current. *J. Neurophysiol.*, *80*, 493–503.
- D’Angelo, E., Nieuwenhuis, T., Maffei, A., Armano, S., Rossi, P., Taglietti, V., Fontana, A., & Naldi, G. (2001). Theta-frequency bursting and resonance in cerebellar granule cells : experimental evidence and modeling of a slow K - dependent mechanism. *J Neurosci*, *21*, 759–770.
- Doiron, B., Chacron, M., Maler, L., Longtin, A., & Bastian, J. (2003). Inhibitory feedback required for network burst responses to communication but not to prey stimuli. *Nature*, *421*, 539–543.
- Doiron, B., Lindner, B., Longtin, A., Maler, L., & Bastian, J. (2004). Oscillatory activity in electrosensory neurons increases with the spatial correlation of the stochastic input stimulus. *Phys Rev Lett*, *93*, 048101.
- Doiron, B., Longtin, A., Turner, R. W., & Maler, L. (2001). Model of gamma frequency burst discharge generated by conditional backpropagation. *J Neurophysiol*, *86*, 1523–1545.
- Gerstner, W., & Kistler, W. (2002). *Spiking neuron models- single neurons, population, plasticity*. Cambridge Univ. Press.

- Gilson, M., Burkitt, A., Grayden, D., Thomas, D., & van Hemmen, J. (2009). Emergence of network structure due to spike-timing-dependent plasticity in recurrent neuronal networks IV: structuring synaptic pathways among recurrent connections. *Biol. Cybern.*, *101*, 427–444.
- Gussin, D., Benda, J., & Maler, L. (2007). Limits of linear rate coding of dynamic stimuli by electroreceptor afferents. *J Neurophysiol*, *97*, 2917–2929.
- Harvey-Girard, E., Lewis, J., & Maler, L. (2010). Burst-induced anti-Hebbian depression acts through short-term synaptic dynamics to cancel redundant sensory signals. *J Neurosci*, *30*, 6152–6169.
- Haykin, S., & Chen, Z. (2005). The cocktail party problem. *Neural Comp*, *17*, 1875–1902.
- Jorntell, H., & Ekerot, C. (2006). Properties of somatosensory synaptic integration in cerebellar granule cells in vivo. *J. Neurosci.*, *26*, 11786–11797.
- Kepecs, A., van Rossum, M. C., Song, S., & Tegner, J. (2002). Spike-timing-dependent plasticity: common themes and divergent vistas. *Biol Cybern*, *87*, 446–458.
- Laing, C. R., & Longtin, A. (2002). A two-variable model of somatic dendritic interactions in a bursting neuron. *Bull Math Biol*, *64*, 829–860.
- Longtin, A., & Chialvo, D. (1998). Stochastic and deterministic resonances in excitable systems. *Phys. Rev. Lett.*, *81*, 4012.
- Marsat, G., Longtin, A., & Maler, L. (2012). Cellular and circuit properties supporting different sensory coding strategies in electric fish and other systems. *Curr. Opin. Neurobiol.*, (p. in press).
- Marsat, G., & Maler, L. (2010). Neural heterogeneity and efficient population codes for communication signals. *J. Neurophysiol.*, *104*, 2543–2555.
- Marsat, G., Proville, R. D., & Maler, L. (2009). Transient signals trigger synchronous bursts in an identified population of neurons. *J Neurophysiol*, (pp. 714–723).
- Nelson, M., Xu, Z., & Payne, J. (1997). Characterization and modeling of P-type electrosensory afferent responses to amplitude modulations in a wave-type electric fish. *J. Comp. Physiol. A*, *181*, 532–544.
- Nelson, M. E., & Maciver, M. A. (1999). Prey capture in the weakly electric fish *Apteronotus albifrons*: sensory acquisition strategies and electrosensory consequences. *J Exp Biol*, *202*, 1195–1203.
- Noonan, L., Doiron, B., Laing, C., Longtin, A., & Turner, R. W. (2003). A dynamic dendritic refractory period regulates burst discharge in the electrosensory lobe of weakly electric fish. *Neurosci Res*, *23*, 1524–1534.
- Oswald, A.-M. M., Chacron, M. J., Doiron, B., Bastian, J., & Maler, L. (2004). Parallel processing of sensory input by bursts and isolated spikes. *J Neurosci*, *24*, 4351–4362.

- Rancz, E., Ishikawa, T., Duguid, I., Chadderton, P., Mahon, S., & Häusser, M. (2007). High-fidelity transmission of sensory information by single cerebellar mossy fibre boutons. *Nature*, *450*, 1245–1248.
- Requarth, T., & Sawtell, N. (2011). Neural mechanisms for filtering self-generated sensory signals in cerebellum-like circuits. *Curr. Opin. Neurobiol.*, *21*, 602–608.
- Roberts, P., & Bell, C. (2000). Computational consequences of temporally asymmetric learning rules: II. sensory image cancellation. *J. Comput. Neurosci.*, *9*, 67–83.
- Roberts, P. D., & Portfors, C. V. (2008). Design principles of sensory processing in cerebellum-like structures. Early stage processing of electrosensory and auditory objects. *Biol Cyber*, *98*, 491–507.
- Sas, E., & Maler, L. (1987). The organization of afferent input to the caudal lobe of the cerebellum of the gymnotid fish *Apteronotus leptorhynchus*. *Anat and Embryol*, *177*, 55–79.
- Saunders, J., & Bastian, J. (1984). The physiology and morphology of two types of electrosensory neurons in the weakly electric fish *Apteronotus leptorhynchus*. *J Comp Physiol A*, *154*, 199–209.
- Sawtell, N. B., & Williams, A. (2008). Transformations of electrosensory encoding associated with an adaptive filter. *J Neurosci*, *28*, 1598–1612.
- Sutherland, C., Doiron, B., & Longtin, A. (2009). Feedback-induced gain control in stochastic spiking networks. *Biol Cybern*, *100*, 475–489.
- Tuckwell, H. C. (1988). *Introduction to Theoretical Neurobiology* volume 2. Cambridge, Mass.: Cambridge university press.
- Vorobyov, S., Cichocki, A., & Bodyanskiy, Y. (2001). Adaptive cancellation for multisensory-signals. *Fluct. Noise Lett.*, *1*, R13–R24.