To appear in *Neural Computation* (accepted August 2010, in press)

Does High Firing Irregularity Enhance Learning?

Chris Christodoulou

cchrist@cs.ucy.ac.cy

Aristodemos Cleanthous

aris@cs.ucy.ac.cy

Department of Computer Science, University of Cyprus, 75 Kallipoleos Avenue, P.O. Box 20537, 1678 Nicosia, Cyprus

Keywords: Partial somatic reset, high firing irregularity, reward-modulated Spike Time-Dependent Plasticity

Abstract

In this paper, we demonstrate that the high firing irregularity produced by the leaky integrate-and-fire neuron with the partial somatic reset mechanism, which has been shown to be the most likely candidate to reflect the mechanism used in the brain for reproducing the highly irregular cortical neuron firing at high rates (Bugmann, Christodoulou & Taylor, 1997; Christodoulou & Bugmann, 2001), enhances learning. More specifically, it enhances reward-modulated Spike Timing-Dependent Plasticity with eligibility trace when used in spiking neural networks, as shown by the results when tested in the simple benchmark problem of XOR as well as in a complex multiagent setting task.

1 Introduction

After analysing spike train recordings from cortical neurons, Softky and Koch (1992, 1993) demonstrated that these cells *in vivo* fire irregularly at high rates. They also showed that the Leaky Integrate-and-Fire (LIF) neuron model, which temporally integrates excitatory postsynaptic potentials generated by independent stochastic input spike trains, failed in reproducing this observed high firing irregularity. While many methods were proposed to reproduce Softky and Koch's findings (for a brief review, see Christodoulou & Bugmann, 2000, 2001), we have shown that a LIF neuron model with partial somatic reset is a very good candidate for reproducing the observed highly irregular firing at high rates by cortical neurons (Bugmann, Christodoulou & Taylor, 1997; Christodoulou & Bugmann, 2001). In this paper, we are investigating whether the high firing irregularity produced by LIF neurons with the partial somatic reset mechanism, when used in spiking neural networks in the benchmark problem of XOR and in a general-sum game, enhances reward-modulated Spike Timing-Dependent Plasticity (STDP) with eligibility trace (Florian, 2007). More specifically, in the case of the general-sum game, we have a multiagent Reinforcement Learning (RL) task with spiking neural networks as agents in the iterated version of Prisoner's Dilemma (PD) (Rappoport & Chammah, 1965) and we are examining whether its cooperative outcome could be enhanced if the LIF neurons of the networks comprising the agents are equipped with partial reset.

2 Methodology

The XOR problem performs the following mapping between two binary inputs and one binary output: $\{0, 0\} \rightarrow 0$; $\{0, 1\} \rightarrow 1$; $\{1, 0\} \rightarrow 1$; $\{1, 1\} \rightarrow 0$. The setup used for testing the XOR computation was the same as the rate-coded input one of Florian (2007) (see Section 4.2 of that paper), although different input frequency rates were used for the two types of networks (having LIF neurons with or without partial somatic reset) in order to ensure that the output firing rate prior to learning was the same for both. We trained both networks with reward-modulated STDP with eligibility trace (Florian, 2007), where the modulation of standard antisymmetric STDP with a reward

	Cooperate (C)	Defect (D)
Cooperate (C)	R(=4), R(=4)	S(=-3), T(=5)
Defect (D)	T(=5), S(=-3)	P(=-2), P(=-2)

Table 1: Payoff matrix of the Prisoner's Dilemma game with the values used in our experiments. Payoff for the Row player is shown first. R is the "reward" for mutual cooperation. P is the "punishment" for mutual defection. T is the "temptation" for unilateral defection and S is the "sucker's" payoff for unilateral cooperation. The only condition imposed to the payoffs is that they should be ordered such that T>R>P>S.

signal leads to RL. As in Florian (2007), we considered that the networks are able to solve the XOR problem, if at the end of an experiment, the output firing rate for the input pattern $\{1, 1\}$ was lower than the output firing rates for the patterns $\{0, 1\}$ or $\{1, 0\}$. The output firing rate for the input pattern $\{0, 0\}$ was always 0, which resulted from the rate coding of the input patterns (as in Florian, 2007).

In the iterated PD (IPD) the payoffs' structure is such that the agents are required to exploit each other in a way that benefits all agents. This game constitutes a challenging problem, given its contradictory nature and lies in the dynamic environment created by the presence of another learner. In its standard one-shot version, the PD is a game summarised by the payoff matrix of Table 1. There are two players, Row and Column. Each player has the choice of either to "Cooperate"(C) or "Defect" (D). For each pair of choices, the payoffs are displayed in the respective cell of the payoff matrix of Table 1. The "dilemma" faced by the players in any valid payoff structure is that, whatever the other one does, each one of them is better off by defecting than cooperating. The outcome obtained when both defect however is worse for each one of them than the outcome they would have obtained if both had cooperated. In the IPD, an extra rule (2R>T+S) guarantees that the players are not collectively better off by having each player alternate between C and D, thus keeping the CC outcome Pareto optimal (Pareto, 1906). It has to be noted that although the cooperative outcome of the IPD we aim for is a valid Nash equilibrium (Nash, 1950) of the game (contrary to the one shot version), in this task we do not explore how to attain equilibria or best responses to any given strategy, but we focus on achieving mutual cooperation.

For the system we implement two spiking neural networks in a multilayer-type architecture as two players competing in the IPD. An input layer of 60 input neurons receiving a common input of 60 Poisson spike trains grouped in four neural populations is shared by each network. These inputs are fully connected to 60 LIF hidden neurons and two LIF output neurons, randomly chosen to either be excitatory or inhibitory. The equations and values of the parameters used are the same as in Florian (2007). The networks learn simultaneously but separately during each round where each network seeks to maximise its own accumulated reward. The architecture of the system is the same as the one we designed and used in Christodoulou, Banfield and Cleanthous (2010), with the difference being that learning is not based on stochastic synaptic transmission (Seung, 2003), but implemented through modulation of STDP with eligibility trace (Florian, 2007). As in Christodoulou, Banfield and Cleanthous (2010), the input to the system is presented for 500ms (which defines the duration of a learning round during which reinforcement is applied) and encodes the decisions just taken, by the firing rate of four groups of Poisson spike trains. In other words, after round k the outcome of the game (at round k) is fed into the system and the synapses are changed according to it (through reinforcement). Decisions for each network are taken according to the relative activation of its output neurons. In the current study of the IPD, the output firing rate of both systems, with or without the partial somatic reset mechanism in their LIF neurons, was kept the same. This was done by providing greater input frequency to the system comprising of LIF neurons with total reset, in order to compensate for the increased output firing rate in the other system due to the partial reset in its LIF neurons. In addition, a high output firing rate of approximately 100Hz was targeted and achieved for both systems, which is within the high rate bound in which cortical cells *in vivo* fire irregularly as identified by Softky and Koch (1992, 1993). This high rate output firing was necessary, as our aim was to investigate whether the firing irregularity at such high rates, which could be reproduced by the LIF neuron model equipped with the partial somatic reset mechanism (Bugmann, Christodoulou & Taylor, 1997; Christodoulou & Bugmann, 2001), strengthens the IPD's cooperative outcome through learning enhancement. It has to be noted that these necessary output firing arrangements had been made before learning took place, as learning would have modified the output firing rates. The rest of the details regarding the decision encoding of the two networks as well as the

reinforcement administration are exactly the same as in Christodoulou, Banfield and Cleanthous (2010) (see Section 2.4 of that paper). A point to note is that the networks are rewarded according to the game's payoff matrix (see Table 1), which is necessary to contain both positive and negative values (like the chosen ones), since the learning algorithm works with positive and negative reinforcements that are directly applied to the synapses.

The partial somatic reset mechanism works as follows: when the membrane potential $u_i(t)$ surpasses the firing threshold θ, then instead of being reset to the resting potential, it is reset to a level $u_i(t) = \beta \times \theta$, where β is the reset parameter, with a value between 0 and 1.

3 Results and Discussion

During each of the rounds of the IPD multiagent RL task, the two networks of the system configuration described in Section 2 seek to maximise their accumulated payoff by cooperating or defecting at every round of the game. Our general aim is to investigate the capability of each network to learn to cooperate in the IPD and more specifically, whether this capability is enhanced by increasing the firing irregularity of each neuron in the network. Similarly, in the XOR computation we are investigating whether such an increased firing irregularity enhances the ability of the system to solve the problem more efficiently, by increasing the suppression level of the output firing rate for input pattern $\{1, 1\}$, in relation to the firing rates of input patterns $\{0, 1\}$ and $\{1, 0\}$. In particular, for each of the two problems investigated, two simulations were performed, one with the spiking neural networks having LIF neurons with total reset (β = 0) and one with partial reset with β = 0.91; this value of the reset parameter was chosen as it was found to produce the observed high firing irregularity at high rates by cortical neurons (Bugmann, Christodoulou & Taylor, 1997; Christodoulou & Bugmann, 2001). More specifically, in Christodoulou & Bugmann (2001), we showed that with the somatic reset value set at β = 0.91, the firing interspike intervals (ISIs) at high rates are: (i) exponentially distributed and (ii) independent; in addition, in Bugmann, Christodoulou & Taylor (1997), we demonstrated that the coefficient of variation (CV) *vs* mean firing ISI curve with β = 0.91 shows a close similarity, firstly with the experimental one (Softky

and Koch, 1992, 1993) and secondly with the theoretical curve for a random spike train with discrete time steps and a refractory time. Therefore, with the choice of the reset parameter β set to 0.91, the firing ISIs are purely temporally irregular (and there are no bursts, that could increase the firing variability), which fulfills our aim to investigate whether high firing irregularity enhances learning. Thus β = 0.91 is the optimal reset value parameter for our purpose and there is no need to see the performance for other reset value parameters, apart of course for β = 0, i.e., total reset (for comparison), for which the firing ISIs at high rates have very low variability and are close to regularity (Softky and Koch, 1992, 1993; Bugmann, Christodoulou & Taylor, 1997; Christodoulou & Bugmann, 2001).

As it can be seen by the results for the XOR problem (Figure 1a), even though both types of network learned the XOR function, the network with the partial somatic reset mechanism in its LIF neurons performed much better in the task, than the one comprising of LIF neurons with total reset. In particular, the former type of network displayed more qualitative results than the latter, as it managed to consistently suppress more the output firing rate for input pattern $\{1, 1\}$, leading to a bigger difference between the output firing rates for input pattern $\{1, 1\}$ and input patterns $\{0, 1\}$ or $\{1, 0\}$. More specifically, in the network consisting of LIF neurons equipped with partial reset, the suppression of the output firing rate for input pattern $\{1, 1\}$ reached 63% of the average output firing rates for input patterns $\{0, 1\}$ and $\{1, 0\}$, while the respective suppression percentage of the network having LIF neurons with total reset reached only 10%.

The results of both simulations in the IPD multiagent RL task are shown in Figure 1b. The difference in the system's performance is evident. Certainly with both configurations the system learns to cooperate, but when each of the competing networks of the system comprises of LIF neurons equipped with the partial somatic reset mechanism, the accumulated payoff is much higher than when there is total reset after each firing spike; this results from the difference in the cooperative outcome. With the partial reset the two networks learned quickly to reach very strong cooperation in order to maximise their long-term reward and achieved the CC outcome 61% of the time on average. On the contrary, with total reset, learning is not as strong, which is evident by the fact that the system exhibited much less cooperation (39% of the time on average). It has to be noted, that even though there is great variability in performance between

learning episodes with sometimes opposing outcomes, most of the times the system learns how to solve the problem, as reflected in the gradient of the average performance of the results (Figure 1b).

These findings from both investigated tasks suggest that the increased firing irregularity at high rates, which results from the introduction of the partial somatic reset mechanism at every LIF neuron of the XOR network and of the networks of the multiagent system, enhances the learning capability of both systems. This is due to the increased suppression of the output firing rate for input pattern $\{1, 1\}$ in relation to the output firing rates for input patterns $\{0, 1\}$ or $\{1, 0\}$ in the XOR problem and the resulting accumulation of higher cooperative reward in the IPD task. More specifically, this high firing irregularity at high rates enhances reward-modulated STDP with eligibility trace. We believe that this is due to more accurate correlations between pre-synaptic and postsynaptic spike timings and reinforcement signals. If firing is regular, then it is possible for two identical spike pairs to be associated with opposite in sign reinforcement signals, confusing thus the direction of the plasticity for a given synapse. High firing irregularity prevents this unnecessary competition by weakening this possibility and thus preventing a possible corruption of the learning algorithm. In addition, as we described in Section 2, for each of the two problems investigated, the output firing rate was kept the same for both systems, with and without partial somatic reset in every LIF neuron. As expected, the output firing rate was influenced by learning in the duration of the experiments, but not to a great extent and in the same manner for the two systems throughout the simulations. For this reason, we can claim that for both problems investigated, the increased efficiency of the system when every LIF neuron is equipped with the partial somatic reset mechanism, is not due to the increased firing rate, but to the enhanced firing irregularity which caused learning enhancement. From our experiments in both studied tasks, we have also observed that the increased levels of temporal irregularity only have 'positive' effects, because they either increase the speed in a successful learning episode, or reverse a failed learning episode in such a way that it becomes successful. It has to be noted that other variant implementations of RL on spiking neural networks by modulating STDP with a reward signal (apart from Florian, 2007), like for example Izhikevich (2007), Faries and Fairhall (2007) and Legenstein, Pecevski and Maass (2008), could equally well be used for obtaining the results presented in this pa-

per. In general, the use of LIF neurons with the partial somatic reset mechanism is very important, as apart from its precise modelling of the high firing irregularity of cortical neurons at high firing rates (Bugmann, Christodoulou & Taylor, 1997; Christodoulou & Bugmann, 2001), it enhances learning as well. It would be interesting to see whether high firing irregularity enhances learning in Seung's reinforcement of stochastic synaptic transmission (Seung, 2003), as well as in policy gradient-based methods as applied to spiking neurons and networks and result in spike-based formulation of reward-based learning (Xie & Seung, 2004; Pfister, Toyoizumi, Barber & Gerstner, 2006; Baras & Meir, 2007; Vasilaki, Frémaux, Urbanczik, Senn & Gerstner, 2009).

Acknowledgments

We gratefully acknowledge the support of the Cyprus Research Promotion Foundation as well as the European Union Structural Funds for grant PENEK/ENISX/0308/82. We are also grateful to the two anonymous referees for their constructive and stimulating reviews.

References

Baras, D. & Meir, R. (2007). Reinforcement learning, spike-time-dependent plasticity, and the BCM rule. *Neural Computation*, *19*, 2245–2279.

Bugmann, G., Christodoulou, C., & Taylor, J. G. (1997). Role of temporal integration and fluctuation detection in the highly irregular firing of a leaky integrator neuron model with partial reset. *Neural Computation*, *9*, 985–1000.

Christodoulou, C. & Bugmann, G. (2000). Near Poisson-type firing produced by concurrent excitation and inhibition. *Biosystems*, *58*, 41–48.

Christodoulou, C. & Bugmann, G. (2001). Coefficient of variation (CV) vs mean interspike interval (ISI) curves: What do they tell us about the brain. *Neurocomputing*, *38-40*, 1141–1149.

Christodoulou, C., Banfield, G., & Cleanthous, A. (2010). Self-control with spiking

and non-spiking neural networks playing games. *Journal of Physiology - Paris*, *104*, 108–117.

Faries, M. A. & Fairhall, A. L. (2007). Reinforcement learning with modulated spike-timing-dependent synaptic plasticity. *Journal of Neurophysiology*, *98*, 3648–3665.

Florian, R. (2007). Reinforcement learning through modulation of spike-timing-dependent synaptic plasticity. *Neural Computation*, *19*, 1468–1502.

Izhikevich, E. M. (2007). Solving the distal reward problem through linkage of STDP and dopamine signalling. *Cerebral Cortex*, *17*, 2443–2452.

Legenstein, R., Pecevski, D., & Maass, W. (2008). A learning theory for reward-modulated spike-timing-dependent plasticity with application to biofeedback. *PLoS Computational Biology*, *4*(10), e1000180.

Nash, J. (1950). Equilibrium points in n-person games. *Proceedings of the National Academy of Sciences of the United States of America*, *36*, 48–49.

Pareto, V. (1906). *Manuale di economia politica*. Milan: Societa Editrice.

Pfister, J., Toyoizumi, T., Barber, D., & Gerstner, W. (2006). Optimal spike-timing-dependent plasticity for precise action potential firing in supervised learning. *Neural Computation*, *18*, 1318–1348.

Rappoport, A. & Chammah, A. (1965). *Prisoners dilemma: a study in conflict and cooperation*. Ann Arbor, MI: University of Michigan Press.

Seung, H. (2003). Learning in spiking neural networks by reinforcement of stochastic synaptic transmission. *Neuron*, *40*, 1063–1073.

Softky, W. & Koch, C. (1992). Cortical cells should fire regularly, but do not. *Neural Computation*, *4*, 643–646.

Softky, W. & Koch, C. (1993). The highly irregular firing of cortical cells is inconsistent with temporal integration of random EPSPs. *Journal of Neuroscience*, *13*, 334–350.

Vasilaki, E., Frémaux, N., Urbanczik, R., Senn, W., & Gerstner, W. (2009). Spike-based reinforcement learning in continuous state and action space: When policy gradient methods fail. *PLoS Computational Biology*, *5*(12), e1000586.

Xie, X. & Seung, H. (2004). Learning in neural networks by reinforcement of irregular spiking. *Physical Review E*, *69*, 41909.

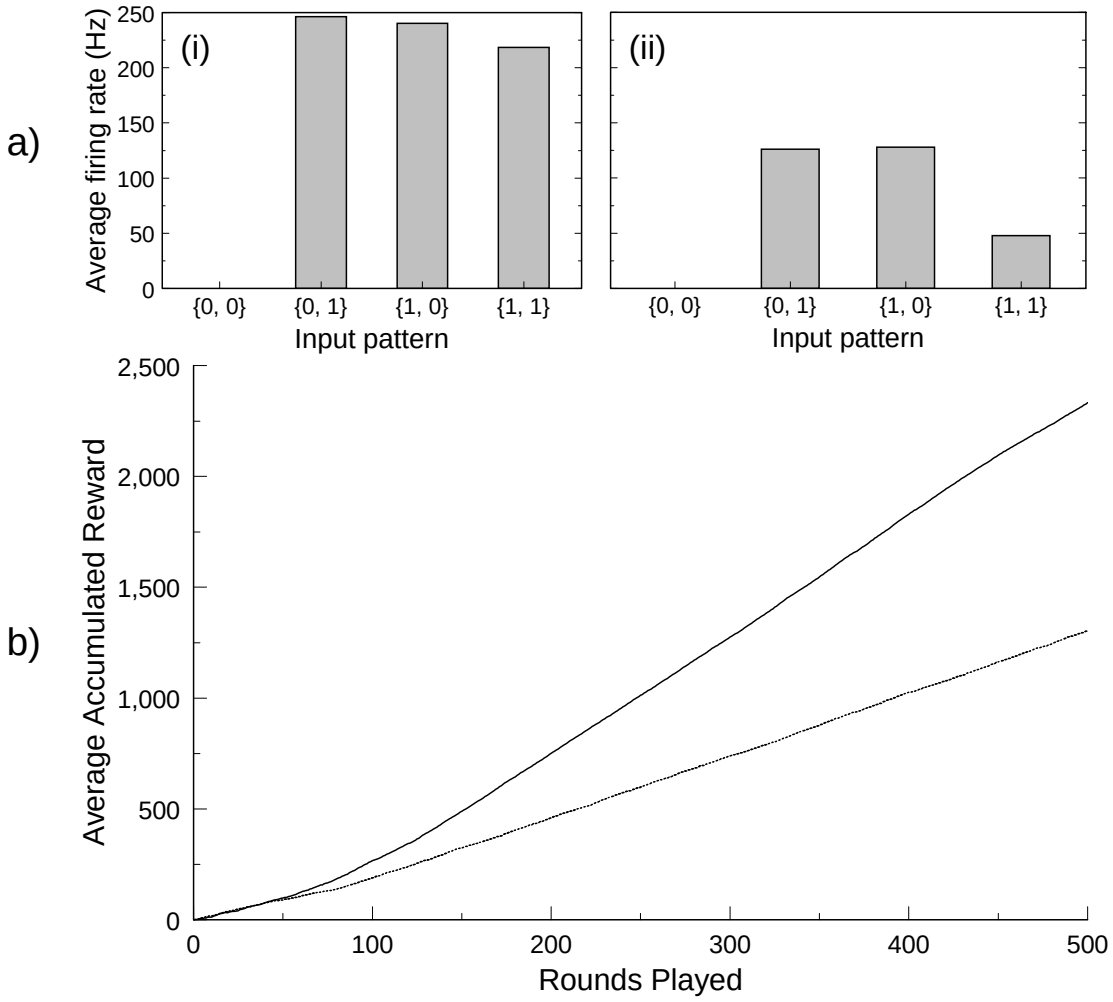

Figure 1: Effect of increased firing irregularity on: (a) learning the XOR computation and (b) learning to cooperate in the IPD. In both problems the networks learn with reward-modulated STDP with eligibility trace (Florian, 2007), whose time constant, τ_z is set to 25ms for all networks and the learning rate γ to 0.00007 for (a) and to 0.0007 for (b) (both found empirically); all other parameter values are as in Florian (2007). (a) Average firing rate of the output neuron after learning, for the four different XOR input patterns with the LIF neurons of the network having either total somatic reset (i), or partial reset at 91% of θ (ii). The recorded results are the averages over 5 experiments. (b) Average Accumulated Reward with the LIF neurons of both networks having either partial somatic reset at 91% of θ (*solid line*) or total reset (*dotted line*). The results recorded are the averages over 10 times of playing the IPD of 500 rounds each.