Short-term synaptic plasticity and heterogeneity in neural systems

J. F. Mejias*, H. J. Kappen†, A. Longtin* and J. J. Torres**

*Dept. of Physics and Centre for Neural Dynamics, University of Ottawa, Canada
†Donders Institute for Brain, Cognition and Behaviour, Radboud Univ. Nijmegen, The Netherlands
**Dept. of Electromagnetism and Physics of the Matter, University of Granada, Spain

Abstract. We review some recent results on neural dynamics and information processing which arise when considering several biophysical factors of interest, in particular, short-term synaptic plasticity and neural heterogeneity. The inclusion of short-term synaptic plasticity leads to enhanced long-term memory capacities, a higher robustness of memory to noise, and irregularity in the duration of the so-called up cortical states. On the other hand, considering some level of neural heterogeneity in neuron models allows neural systems to optimize information transmission in rate coding and temporal coding, two strategies commonly used by neurons to codify information in many brain areas. In all these studies, analytical approximations can be made to explain the underlying dynamics of these neural systems.

Keywords: short-term depression and facilitation, storage capacity, up and down states, heterogeneity, neural coding
PACS: 87.19.La

INTRODUCTION

Theoretical and computational modeling has become a powerful tool to deepen our the understanding of neural systems. This is especially important when dealing with mechanisms or neural circuits that are not easily accessible experimentally, or when experimental data by itself is not enough to provide a clear picture of the phenomena under study. A prominent biophysical mechanism that fits in such a framework is short-term synaptic plasticity, also known as dynamic synapses [1]. This type of fast, activity-dependent variation of the synaptic strength has been shown to have a strong impact on a number of important neural tasks, such as cortical gain control [2], coincidence detection [3, 4], sound localization [5], broadband coding [6], up and down cortical transitions [7, 8], working memory [9] or network storage capacity [10, 11, 12]. There are still a number of open questions, and thus short-term synaptic plasticity deserves further attention from the theoretical and computational point of view.

Another important feature that can be studied in an efficient manner by using theoretical and computational modeling is the effect of cellular heterogeneity on the performance of neural networks. When investigating neural coding strategies used by brain circuits, most theoretical and computational studies do not take into account the intrinsic variability found among actual neurons. In addition, such variability is also usually difficult to control experimentally, making neural heterogeneity a typically underrated factor in neural information processing studies. In recent years, however, a number of studies have suggested a functional role of the intrinsic heterogeneity found in real neu-
ral systems, and in particular its possible influence on network synchronization [13, 14], signal transmission [15, 16], neural coding [17] and sensory processing [18, 19].

In this work, we review some of our recent results on these topics. First, we consider the effect of short-term synaptic plasticity on the storage abilities of attractor neural networks, showing that it leads to enhanced memory capacities [11, 12] and a higher robustness to noise [12]. Then, we consider a simplified model of a neural population to investigate the effect of short-term synaptic plasticity on the irregularity of the duration of up states in cortical networks [8], which has been observed in experiments [20, 21, 22]. Finally, we address the effect of some level of heterogeneity among neurons in a network on the optimization in the information transmission for rate coding and temporal coding, two strategies commonly used by neurons to codify information in many brain areas [17]. In the three studies presented, analytical approximations can be made to explain the underlying dynamics of these neural systems.

**MAXIMUM STORAGE CAPACITY**

We first study the influence of dynamic synapses on the storage abilities of neural networks (for more details, see [11, 12] and references therein). We consider a fully connected network of $N$ binary neurons whose state $s \equiv \{s_i = 0, 1; \forall i = 1, \ldots, N\}$ follows a probabilistic parallel dynamics

$$P[s_i(t + 1) = 1] = \frac{1}{2} \{1 + \tanh[2\beta(h_i(s,t) - \theta_i)]\} \quad \forall i = 1 \ldots N,$$

where $h_i(s,t)$ is the local field or the total input synaptic current to neuron $i$, namely

$$h_i(s,t) = \sum_{j \neq i} \omega_{ij} x_j(t) u_j(t) s_j(t).$$

Also, $\beta \equiv T^{-1}$ is a temperature or noise parameter (i.e., for $\beta \to \infty$ we have a deterministic dynamics), and $\theta_i$ represents the neuron firing threshold. The coefficients $\omega_{ij}$ are fixed synaptic weights, consequence of the slow learning process of $M$ memory patterns of activity. In the following we choose the following learning rule

$$\omega_{ij} = \frac{1}{Nf(1-f)} \sum_{\mu=1}^{M} (\xi^\mu_i - f)(\xi^\mu_j - f),$$

where $\{\xi^\mu_i = 0,1; i = 1 \ldots N\}$ represents the $M$ stored random patterns with mean activity over the patterns $\langle \xi^\mu_i \rangle = f = 1/2$. On the other hand, the variables $x_j$, $F_j$ appearing in $h_i$ describe the short-term depression and facilitation synaptic mechanisms, respectively. We assume that these variables evolve according to the discrete dynamics [23, 24]

$$x_j(t + 1) = x_j(t) + \frac{1-x_j(t)}{\tau_{\text{rec}}} - U_{\text{SE}} F_j(t) x_j(t) s_j(t)$$

$$u_j(t + 1) = u_j(t) + \frac{1-u_j(t)}{\tau_{\text{fac}}} + (1 - U_{\text{SE}} u_j(t)) s_j(t).$$
FIGURE 1. Critical network load in the presence of short-term synaptic plasticity for $U_{SE} = 0.2$. Left: network load vs depression time constant, for different values of the facilitation time constant. Right: critical network load vs facilitation time constant, for different values of the depression time constant (the inset shows the low $\tau_{fac}$ regime, where maxima appear). As one can see, the presence of facilitation allows optimal retrieval abilities in the network while preserving the dynamic nature of synapses. Mean-field predictions (lines) agree with simulations (symbols), which were done with a network of $N = 3000$ neurons.

Here, $\tau_{rec}, \tau_{fac}$ are the time constants for depressing and facilitating processes respectively, and $U_{SE}$ is a parameter related with synaptic unreliability. The original Hopfield model is recovered when $x_i = u_i = 1, \forall i, t$ (i.e. static synapses). By simple inspection of Eqs. (4-5), this corresponds to the case of $\tau_{rec}, \tau_{fac} \ll 1$ which makes $x_j$ and $u_j \forall j$ quickly reach their maximum values, $x_j = u_j = 1 \forall j, t$ (see [11] for a careful explanation of this limit). We also choose

$$\theta_i = \frac{1}{2} \sum_{j \neq i} \omega_{ij}. \quad (6)$$

The network load is defined as $\alpha \equiv M/N$. One can obtain a mean field solution of the network at the limit of zero temperature (see [11] for details) and obtain the following expression for the critical network load $\alpha_c$ (defined as the maximum number of patterns per neuron that the network is able to retrieve), simplified here for clarity purposes:

$$\alpha_c \sim \frac{0.138}{1 + \left( \frac{1 + \gamma' - \gamma}{\gamma} \right)^2}, \quad (7)$$

where $\gamma \equiv U_{SE} \tau_{rec}$ and $\gamma' \equiv \frac{1 + \tau_{fac}}{1 + U_{SE} \tau_{fac}}$. We can employ such a mean field solution to compare with the numerical simulations of the model and study the effect of the time constants $\tau_{rec}, \tau_{fac}$ on the critical network load. As Fig. 1 shows, the presence of purely depressing synapses (that is, $\tau_{fac} = 0$) leads to low critical network loads; this effect is stronger for larger $\tau_{rec}$ (see also [24]). On the other hand, the presence of a certain level of facilitation allows an optimal network load (and therefore optimal retrieval abilities) for certain finite nonzero values of $\tau_{rec}, \tau_{fac}$. This is highly desirable from both computational and biophysical points of view, since one would want to have a neural
network with good retrieval abilities while preserving the dynamic nature of synapses, which endow it with further information processing and coding capabilities.

One can also extend the mean-field solution found in [11] to the case of nonzero temperatures, in order to study the behavior of the network in more general conditions (see [12] for details). In particular, we can compute the phase diagrams of the model as a function of the network load and temperature, and then investigate the effect of short-term plasticity on these diagrams. As the left panel of Fig. 2 shows, the presence of short-term facilitation increases the area of the memory phase, even in the presence of short-term depression [12]. The effect on the critical temperature $T_c$, which is the maximum temperature allowing good retrieval abilities and corresponds to the $\alpha \to 0$ case, is worth mentioning. We can observe clearly in the right panel of Fig. 2 that increasing $\tau_{\text{fac}}$ leads to large values of $T_c$, for different values of the depression time constant. Such dependence can be found analytically [12], and it is $T_c = \gamma'/(1 + \gamma'\gamma)$. These large $T_c$ values comfortably surpasses the critical temperature of the standard Hopfield model (marked as a dashed line in the figure), indicating that the presence of facilitation allows the network to perform optimally in retrieval tasks even in strong noise conditions.

**IRREGULARITY OF CORTICAL UP STATES**

We have seen so far that, from a purely theoretical standpoint, short-term synaptic plasticity, and in particular short-term facilitation, may be highly beneficial for a neural network to improve information retrieval properties. However, short-term depression by itself is also crucial to explain certain features observed in actual neural systems. A good example of this is the spontaneous transitions between activity states observed in cortical areas in the brain, a phenomenon which is referred to as up and down transitions. Such behavior may provide a framework for neural computations [25], and could also coordinate some sleep rhythms into a coherent rhythmic oscillatory behavior in cortical
and thalamocortical areas [26, 27, 28].
A prominent feature of up and down transitions is that, for certain experimental conditions, the times between transitions seem to be highly irregular, and in particular the duration of up states is found to range from a scale of milliseconds to seconds [20, 21, 22]. While previous modeling studies (which consider up and down transitions as a phenomenon induced by synaptic depression) are not able to explain such irregularity [7], a combination of short-term depression and other biophysical factors could still be able to explain such erratic behavior. Here we briefly discuss a simplified model of a neural population, with short-term depression and synaptic noise, which is able to explain the irregularity found in the duration of up states (for more details, see [8]).

We assume that the dynamics of the firing rate of the neural population is described by

\[
\frac{d\nu(t)}{dt} = -\nu(t) + F[Jx(t)\nu(t) - \theta] + \zeta(t),
\]

where \(\nu(t)\) is the mean firing rate or activity of the neural population, \(J\) is the synaptic coupling strength in absence of short-term depression, and \(\theta\) is a firing threshold. The variable \(\zeta(t)\) is a Gaussian white noise of zero mean and standard deviation \(\delta\), which takes into account the inner stochasticity of neurons. The term \(F[z] \equiv \frac{\nu_{\text{max}}}{2}(1 + \tanh(z))\) is a sigmoidal function, which sets the up and down activity levels to \(\nu = \nu_{\text{max}}\) and \(\nu = 0\), respectively.

The variable \(x(t)\) takes into account the synaptic variations due to short-term depression, and evolves according to

\[
\frac{dx(t)}{dt} = \frac{1 - x(t)}{\tau_{\text{rec}}} - U_{SE} x(t)\nu(t) + \frac{D}{\tau_{\text{rec}}} \xi(t),
\]

where \(\tau_{\text{rec}}\) is again the depression time constant, and \(U_{SE}\) is a parameter related with the synaptic unreliability (we do not consider short-term facilitation here). The last term (with \(\xi(t)\) being a Gaussian white noise of zero mean and unitary variance, and \(D\) a constant representing the strength of the noise) takes into account any source of uncontrolled noise that is not included in \(\zeta(t)\) and that could directly influence the synaptic strength, such as fluctuations in neurotransmitter release or receptor unreliability [29, 30].

A typical temporal evolution of this model is shown in Fig. 3A, while Fig. 3B corresponds to the histogram of population activity values, which reflects the bimodal nature found in experiments. The activity of up and down states can be easily identified as 5 spikes/s and 0.5 spikes/s, respectively.

One can develop a theoretical estimation of the probability distribution \(P(T)\) of a certain up state duration \(T\) (such as the one displayed in Fig. 3A). Briefly, by carefully analyzing the conditions for the existence of a double well in the dynamics (8), one can find that only a limited window of values of \(x(t)\) is compatible with the existence of such a double well. When \(x(t)\) is above (below) this window, the double well turns into a single well centered at the up (down) activity state, respectively. If the synaptic noise is strong enough, the variable \(x(t)\) will be constantly pulled away from this window, and as a result the synaptic strength will drive the transitions between up and down states. In practice, the double well window for \(x(t)\) will be small and centered on its mean value
FIGURE 3. (A) Transitions between up and down states in a neural population with depressing synapses. In the model, the transition between the two possible activity levels is driven by the average strength of the synapses at a given time, namely $x(t)$. Population activity (top) as well as averaged synaptic strength (bottom) are shown, with the mean value of $x(t)$ denoted by a dashed line. (B) Histogram of the population firing rate, displaying the bistable nature of the population dynamics. (C) Distribution of duration times of the up state, for different levels of synaptic noise. The slope $\lambda_{\text{exp}} = 1.43$ corresponds to the value found in in vitro experiments (see main text for details). Parameters are $J = 1.1$, $U_{SE} = 0.6$, $\tau_{\text{rec}} = 1000 \text{ ms}$, $\delta = 0.3$ and $v_{\text{max}} = 5 \text{ spikes/s}$.

$\lambda_{\text{exp}}$, One can see this effect in Fig. 3A, where large deviations of $x(t)$ from its mean value (dashed line) drive the population activity towards up or down states.

In this context, the presence of temporal autocorrelations in $x(t)$ (reflected by the short-term depression time constant $\tau_{\text{rec}}$) will increase the duration of the excursions of $x(t)$ out of the double well window, which in turn will imply an increment in the duration of the up states. Thus, calculating the distribution of ruin times of $x(t)$ (that is, the mean duration of the excursions of $x(t)$ away from its mean value) will give us the distribution of up state durations, which results in $P(T) \sim T^{-\lambda}$ with $\lambda = 1.5$.

The numerical simulations of our model indicate, as can be seen in Fig. 3C, that the distribution of up state durations $P(T)$ tends to a power-law distribution with slope $\sim 1.5$ for sufficiently large values of the synaptic noise $D$, as our theoretical estimations predict. The figure also shows, for a direct comparison, the slope of $\lambda_{\text{exp}} \sim 1.43$ found in in vitro experiments [21], in a very good agreement with our theoretical and numerical predictions.
HETEROGENEITY AND NEURAL CODING

In this last section, we focus on the effect of neural heterogeneity on the coding properties of neural networks (a detailed description of this section can be found in [17]). We start by considering a fully connected network of $N$ excitatory neurons, where the dynamics of each neuron $i$ is described by

$$\tau_m \frac{dV_i(t)}{dt} = -V_i(t) + S(t) + \mu + \sqrt{\tau_m} \sigma \xi_i(t) + \frac{\tau_m}{N} \sum_j \sum_k J \delta(t - t^k_{ij}),$$

(10)

where $\tau_m$ is the neuron membrane time constant, $V_i$ is its membrane potential, $S(t)$ is an external input signal to be determined, $\mu$ is a constant input bias, $\xi_i(t)$ is a gaussian white noise of zero mean and unitary variance, $\sigma$ is the noise strength, $J$ is the synaptic coupling strength, and the $k$-th spike from neuron $j$ arrives at neuron $i$ at $t^k_{ij}$; the effect of this spike on the neuron is modeled as a delta-like pulse. Each neuron $i$ is assumed to fire an action potential (AP) every time $V_i$ reaches a certain firing threshold, and after that the membrane potential is reset to $V_r$ for a time period $\tau_{ref}$. In addition, we assume here that each neuron $i$ has a firing threshold $\theta_i$ which is randomly distributed following a gaussian profile $P(\theta)$ with mean $\bar{\theta}$ and standard deviation $w$. Such heterogeneity reflects some of the variability in the individual excitability properties of neurons found in actual neural systems. The network will be homogeneous for $w = 0$, when the distribution of thresholds becomes a delta centered at $\bar{\theta}$.

We now consider that the external signal $S(t)$ is a weak, low-frequency sinusoidal signal. Assuming a homogeneous network (that is, $w = 0$) that remains in an asynchronous state (see [31]), the signal is able to slowly modulate the mean firing rate of the network (see Fig. 4A). Such modulation of the firing rate, which captures and transmits the information of the signal, is referred to as rate coding [32, 33]. In order to investigate the effects of neural heterogeneity on information transmission and neural coding, we now assume a certain level of heterogeneity in the network (so $w > 0$) and compute the modulation of the mean firing rate due to $S(t)$ in this case. We observe, both analytically and numerically, that the strength of the modulation depends on the level of heterogeneity in a non-trivial way (Fig. 4B). In particular, we found a non-monotonic behavior of the modulation (measured as an input-output correlation) with $w$, suggesting that there is a certain neural heterogeneity level which optimizes information transmission under rate coding. This optimization is found to be caused by the nonlinear effect of heterogeneity on the baseline mean firing rate of the network [17].

On the other hand, if we assume that the working point of our homogeneous network ($w = 0$) is close to the stability line of the system, a small external perturbation may be able to destabilize the network and make the neurons synchronize briefly, producing a population spike [31, 34]. Networks can use this high sensitivity to small external perturbations to process well time-located incoming signals, as seen in Fig. 4C. Such a detection strategy, which strongly relies on the generation of sharp responses precisely located in time, is known as temporal coding [32, 33]. When considering the effect of neural heterogeneity in information transmission under temporal coding (for which we use here the positive predictive value, PPV, a widely used measure [35]), we observe a nonlinear dependence as well (see Fig. 4D). More precisely, we find a non-monotonic
FIGURE 4. Effect of neural heterogeneity on the coding properties of spiking neural networks. (A) Modulation of the network mean firing rate (black) due to a weak sinusoidal signal (grey); the network is in an asynchronous working regime and $w = 0$. (B) Signal transmission, measured as the input-output correlation of the system, is optimized for a certain nonzero level of neural heterogeneity under rate coding. Theoretical estimation (line) agrees with numerical simulations (points). (C) Detection of sharp inputs (triangles) by the network through population-spike generation. The network is close to the stability line, and $w = 0$. (D) Signal transmission, measured as the PPV, is optimized for a certain nonzero level of neural heterogeneity under temporal coding. Parameter values are $N = 1500$ (for simulations), $\tau_m = 20\, ms$, $\mu = 14\, mV$, $V_r = 10\, mV$, $\tau_{ref} = 5\, ms$, $\bar{\theta} = 20\, mV$ and $\sigma = 3\, mV$. Coupling strength $J$ is $10\, mV$ (A, B) or $20\, mV$ (C, D).

shape of the PPV with $w$, suggesting that a certain neural heterogeneity value is able to optimize the detection of signals under temporal coding. This phenomenon is due to a diversity-induced synchronization of the neurons which leads to a resonance-like behavior of the system [15, 16, 17].

CONCLUSION

In this work, we have reviewed some recent results concerning the role of several factors, such as short-term synaptic plasticity or neural heterogeneity, on the dynamics of neural networks. In the case of short-term synaptic plasticity, we have shown that short-term facilitation has a positive impact (with respect to purely depressing synapses) in the retrieval abilities of attractor neural networks [11]. In particular, the presence of facilitation allows to have a neural network with good retrieval abilities while preserving the natural fast dynamics of synaptic weights, convenient for information processing and coding. Furthermore, the storage properties of attractor neural networks are improved
with facilitation also in high-noise conditions, yielding a larger area of good memory retrieval and higher critical temperatures [12]. This suggests that facilitation may have an important role in helping actual neural circuits to access and maintain previously stored information in the presence of strong noise, which would occur, for instance, in working memory tasks [9].

Short-term depression, on the other hand, has been found to have a major impact on several tasks, such as gain control [2] or sound localization [5]. As we reviewed here, the combination of short-term synaptic depression and synaptic noise may be responsible for the irregularity of the duration of up states [8]. Such irregularity has been observed in several experimental conditions [20, 21, 22], although there is also experimental evidence of a much lower irregularity in different conditions [22, 36, 37]. Therefore, further theoretical and experimental work is needed to understand the origin of the irregularity of up state duration, and why it is absent in some situations. This absence could involve, for instance, underlying mechanisms modulating the level of short-term depression or the strength of synaptic noise.

Finally, we have analyzed the role of intrinsic neural heterogeneity on the coding properties of spiking neural networks. Contrary to what is traditionally assumed, heterogeneity on neural systems does not only allow networks to process information properly, but a certain level of such heterogeneity may even boost the abilities of the network to process and transmit information under several neural strategies, namely, rate coding and temporal coding. Such results are found to be robust in more realistic structures, such as in sparsely connected networks of excitatory and inhibitory neurons [17]. The way in which short-term synaptic plasticity and neural heterogeneity could interact in the processing of information constitutes an open question which is currently under study.

ACKNOWLEDGMENTS

This work was supported by Ministerio de Ciencia y Tecnología and Junta de Andalucía of Spain, and NSERC Canada. We thank J. Marro for useful discussions.

REFERENCES