

# BIO-465 – Project

## A network model of cortical surround suppression

Etienne Objois, Niels Vadot  
EPFL, Switzerland

**Abstract**—We explore neuronal population models and use them to model cortical surround suppression. Cortical surround suppression is a phenomenon where the presence of nearby neighbours causes a neuron to have a lower activity level than if it were alone. We also explore the orientation tuning phenomenon, where the surround suppression is more or less strong, depending on the overlap between the external stimulus and the receptive fields. The source code is available at [1].

### I. RATE MODELS OF NEURONAL POPULATIONS

A neuronal population is composed of multiple neurons, which can each be individually modeled. In the simple model from EQ. (1), a neuron is described by the voltage across its membrane  $h(t)$ , resistivity  $R$  to input currents  $I(t) = I_{\text{ext}}(t) + I_{\text{network}}(t)$ , a relaxation time constant  $\tau$ , and an activity level  $A(t)$  defined in terms of a filter  $\alpha(s)$  and gain function  $F(h)$  in the equation  $A(t) = \int_0^\infty \alpha(s) F(h(t-s)) ds$ .

$$\tau \dot{h}(t) = -h(t) + RI(t) \quad (1)$$

In a population, every neuron is modeled individually according to EQ. (1), and attributed an index  $k$ . We furthermore model the network currents linearly as in EQ. (2), by introducing a weight parameter  $W_{kn}$ , which describes the influence of neuron  $n$  on neuron  $k$ .

$$I_{\text{network},k}(t) = \sum_n W_{kn} A_n(t) \quad (2)$$

A further simplification can be made by letting  $\alpha(s)$  be a very sharp filter, approximatively a Dirac delta. Then  $A(t) \approx F(h(t))$ .

We can rewrite the entire system as one equation in matrix form, where  $\mathbf{h} = (h_1, \dots, h_n)$ ,  $\mathbf{R} = (R_1, \dots, R_n)$ , etc., and  $\odot$  is the elementwise product :

$$\tau \odot \dot{\mathbf{h}} = -\mathbf{h} + \mathbf{R} \odot (\mathbf{W}\mathbf{F}(\mathbf{h})) + \mathbf{R} \odot \mathbf{I}_{\text{ext}} \quad (3)$$

EQ. (3) models a population of neurons, but can also be interpreted to model an ensemble of populations. This is assumed from now on, and a more detailed proof and renormalization is given in subsection III-C.

### II. FIXED VS. RECURRENT INHIBITION

#### A. Excitatory population with self-coupling

We model a single population according to EQ. (3), and assume constant input current  $I_{\text{ext}}$ . The gain function is defined as  $F(h) = \sigma(h) = \frac{1}{2}(1 + \tanh(h))$ .

The self-coupling appears in the diagonal terms of  $\mathbf{W}$ , and in this case there is only self-coupling, since  $\mathbf{W}$  reduces to a scalar.

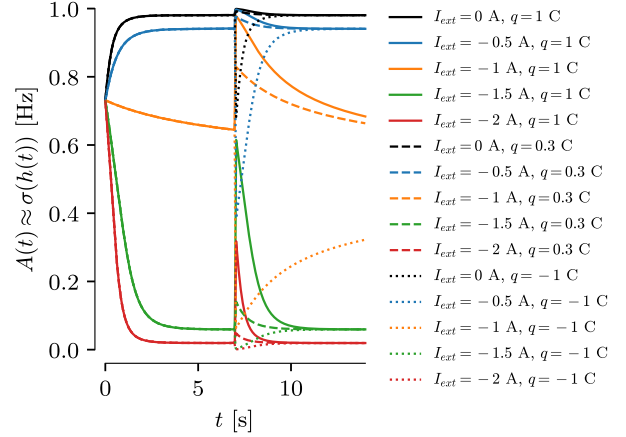


FIGURE 1: Single population with constant input and a delta spike of total charge  $q$  at  $t = 7$ . Parameters  $W = 2$ ,  $\tau = 0.6$ ,  $R = 1$ ,  $h(0) = 0.5$ .

We vary the constant input current in FIG. 1 and study  $t < 7$  (before the delta spike). The simulation is started at  $h(0) \neq 0$  in order to break symmetry of the gain function. For all currents, the activity levels out over time. We can observe three regimes. (1) At  $I_{\text{ext}} \lesssim -1$  A, there is too much negative current (i.e. too many negative charges arrive too fast, which depolarizes the population) and the self-coupling is not enough to maintain activity levels, which decay to nearly zero. (2) At  $I_{\text{ext}} \gtrsim -1$  A, the opposite happens, and self-coupling manages to drive the activity levels to nearly one. (3) At  $I_{\text{ext}} \approx -1$  A, the external current and self-coupling nearly cancel out, leaving the activity somewhere between zero and one.

The introduction of a delta spike at  $t = 7$  makes the activity levels jump, which then recover and tend back to the same equilibrium as before the spike. For a same total charge, unsaturated neurons ( $0 \lesssim A \lesssim 1$ ) jump a larger amplitude and take more time to recover than saturated neurons ( $A \approx 0$  or  $A \approx 1$ ). The jump is asymmetric with respect to charge sign : if  $A > \frac{1}{2}$ , then for  $q < 0$  the perturbation is greater than for  $q > 0$  (and vice-versa). For total charges of the same sign, the size of the jump is largest for the largest absolute charge.

In order to get insight into the behavior of the coupling strength, FIG. 2 shows a 1D phase plot of the differential equation. Fixed points solutions to  $\dot{h} = 0$ .  $I_{\text{ext}}$  controls the vertical offset of the curves but more interestingly  $W$  controls the appearance of two local extrema, and there-

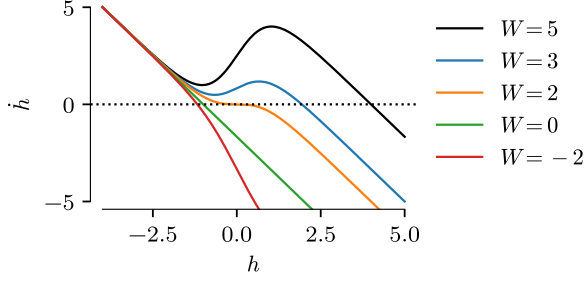


FIGURE 2: 1D phase plot of a single population. Parameters  $\tau = 0.6, R = 1, I_{\text{ext}} = -1$ . Critical values are  $W_{\text{crit}} = 2$  and  $I_{\text{crit}} = -1$ .

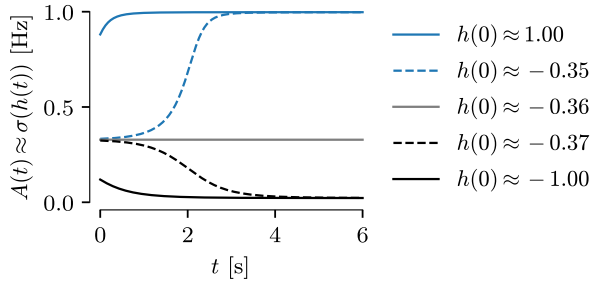


FIGURE 3: Simulation of a single population with three fixed points. Parameters  $W = 5, R = 1, I_{\text{ext}} = -2$ .

fore the possibility of two other solutions. This happens when the curvature changes sign, and critically when the derivative is null. Solving  $\partial_h^2 \dot{h} = 0$  and  $\partial_h \dot{h} = 0$  for  $h$  and  $W$ , we find  $h_{\text{crit}} = 0$  and  $W_{\text{crit}} = 2/R$ . When  $W < W_{\text{crit}}$ , only one stable solution exists. When  $W > W_{\text{crit}}$ , three solutions can exist (two stable, and one unstable at  $h = 0$ ) depending on the value of  $I_{\text{ext}}$ .

If we furthermore impose  $\dot{h} = 0$  at the critical value, then we get  $I_{\text{crit}} = -1/R$ . When  $I < I_{\text{crit}}$ , the three solutions can exist (provided the current is not too negative), whereas when  $I > I_{\text{crit}}$  only one solution exists.

Intuitively, the current has to be depolarizing enough but the self-coupling strong enough to compensate, and create a second solution. FIG. 3 demonstrates a set of parameters for which there are two stable solutions, and one unstable solution.

### B. Linear rate networks

We specialize EQ. (3) in the case of a linear rate network, that is  $F(h) = \text{Id}(h) = h$ . Assuming external current to be independent of time, this equation is of the form  $\dot{\mathbf{h}} = \Lambda \mathbf{h} + \mathbf{b}$ , where  $\Lambda = (1/\tau) \odot (R \odot W - \text{Id})$  and  $\mathbf{b} = (1/\tau) \odot (R \odot I_{\text{ext}})$ . This can be solved analytically :

$$\begin{aligned} \dot{\mathbf{h}} &= \Lambda \mathbf{h} + \mathbf{b}, \text{ changing variables } \mathbf{y} = \Lambda \mathbf{h} + \mathbf{b} \\ \implies \dot{\mathbf{y}} &= \Lambda \mathbf{y} \implies \dot{\mathbf{y}} = \exp(\Lambda t) \mathbf{y}_0 \end{aligned}$$

We can change back variables, denoting the Moore-

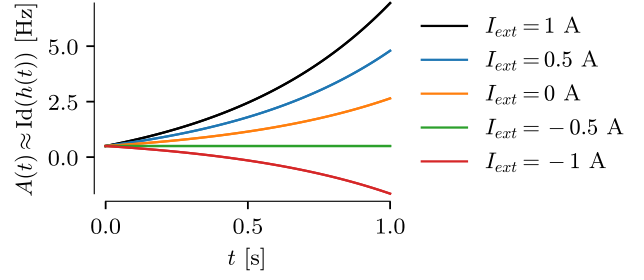


FIGURE 4: Simulation of a linear single population. Parameters  $W = 2, R = 1, \tau = 0.6, h(0) = 0.5$ .

Penrose inverse  $\Lambda^+$ , and get the general solution EQ. (4).

$$\mathbf{h}(t) = \begin{cases} \Lambda^+(\exp(\Lambda t)(\Lambda \mathbf{h}_0 + \mathbf{b}) - \mathbf{b}) & \text{if } \Lambda \neq 0 \\ \mathbf{b}t & \text{if } \Lambda = 0 \end{cases} \quad (4)$$

The stability of the solution (in the case  $\Lambda \neq 0$ ) becomes clear by rewriting  $\Lambda \mathbf{h}(t) = \exp(\Lambda t) \mathbf{v} - \mathbf{b}$ , where  $\mathbf{v} = \Lambda \mathbf{h}_0 + \mathbf{b}$ . Recall that the eigenspace of  $A$  is invariant under matrix exponentiation.

Components of  $\mathbf{v}$  that lie in an eigenspace associated to  $\lambda < 0$  decay exponentially, diverge if  $\lambda > 0$ , and behave linearly if  $\lambda = 0$ . This is demonstrated in FIG. 4, where depending on  $I_{\text{ext}}$ , the activity either stays constant (linear behavior) or diverges to  $\pm\infty$  exponentially.

In a linear rate network, the presence of a inhibitory current is not sufficient to ensure stabilization of a single population, whereas it was sufficient for a nonlinear rate network (in fact, in the case  $F(h) = \sigma(h)$  the activity is bounded in  $[0, 1]$  so the network is always stable.).

### C. Inhibition-stabilized network (ISN)

We add another population to the network. The weights of the outbound currents (second column of  $W$ ) of the new population are negative, such that it can be seen as the inhibitory (I) population. The original population has positive outbound weights (first column of  $W$ ), becoming the excitatory (E) population.

To study stability of the network, we compute the eigenvalues of  $\Lambda$  from subsection II-B, and find a condition such that they are all negative.

$$\Lambda = \begin{pmatrix} \frac{W_{00}-1}{\tau_0} & \frac{W_{01}}{\tau_1} \\ \frac{W_{10}}{\tau_1} & \frac{W_{11}-1}{\tau_1} \end{pmatrix}.$$

We find that both eigenvalues change sign at

$$W_{00}^* = \frac{W_{01}W_{10} + W_{11} - 1}{W_{11} - 1}.$$

Notice that since  $W_{11} < 0$ , we have  $W_{11} \neq 1$  and a stability threshold exists for any  $W_{10}, W_{01}, W_{11}$ .

In this case, we can substitute  $W_{10} = 4, W_{01} = -4, W_{11} = -7$  and find the system is stable when  $W_{00} < W_{00}^* = 3$ . It diverges for  $W_{00} > W_{00}^*$  and evolves linearly when  $W_{00} = W_{00}^*$  (both eigenvalues are null with the same condition).

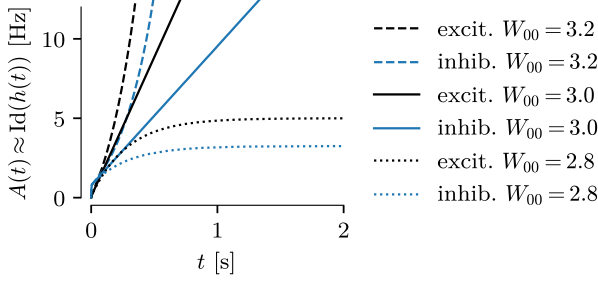


FIGURE 5: Simulation of a linear ISN. Parameters  $W_{10} = 4, W_{01} = -4, W_{11} = -7, R = 1, \tau = (0.06, 0.012), \mathbf{h}(0) = (0, 0), \mathbf{I}_{\text{ext}} = (4.0, 6.0)$ .

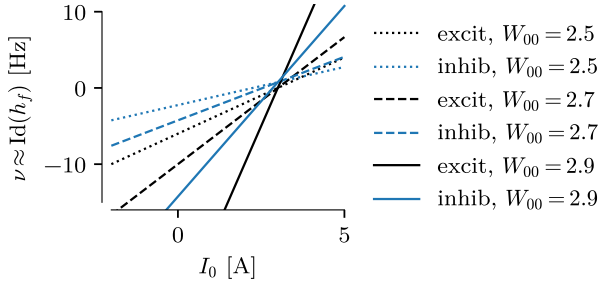


FIGURE 6: Simulation of a linear ISN. Parameters  $W_{10} = 4, W_{01} = -4, W_{11} = -7, R = 1, \tau = (0.06, 0.012), \mathbf{I}_{\text{ext}} = (4.0, 6.0)$ .

The numerical simulations (setting  $W_{10} = 4, I_0 = 4.0, I_1 = 6.0$ ) shown in FIG. 5 confirm the analytical derivations. In the converging case  $W_{00} = 2.8 < W_{00}^* = 3.0$ , the excitatory activity converges to  $A = 5$  Hz, and the inhibitory activity converges to  $A = 3.25$  Hz.

We denote  $\mathbf{h}_f = \Lambda^+ \mathbf{b}$  (see subsection A) the value to which  $\mathbf{h}$  converges. It depends on  $W, \mathbf{I}_{\text{ext}}$ , and more surprisingly  $\tau$ . In the case of one population, there was no "competition" and the system can converge at any speed. In the case of two populations, the competition means that now the rate of convergence matters.  $\mathbf{h}_f$  is of the form, where cst terms are functions of  $W$  and  $\tau$ :

$$\mathbf{h}_f = \begin{pmatrix} \text{cst} \cdot I_0 + \text{cst} \cdot I_1 \\ \text{cst} \cdot I_0 + \text{cst} \cdot I_1 \end{pmatrix}.$$

The f-I curve shown in FIG. 6 is linear, as expected from the form of  $\mathbf{h}_f$ . Stronger excitatory self-couplings of the E population are associated to a steeper line. Compared to the linear excitatory-only network which (nearly) always diverges exponentially, the addition of an inhibitory population allows for the existence of stable regimes. This shows that constant inhibitory current is not enough to stabilize the system, it needs to scale with the activity of the excitatory population, as it does here.

This linear analysis extends to the sigmoidal gain function  $F(h) = \sigma(h)$  in the case of a small potential  $h$  (after some rescaling), since  $\sigma(h) = \frac{1}{2} + \frac{h}{2} + \mathcal{O}(h^3) \approx \text{Id}(\frac{h+1}{2})$ .

### III. MODELING SURROUND SUPPRESSION

#### A. Network mechanisms of surround suppression

#### B. Orientation tuning of surround suppression

#### C. Surround suppression in networks with bio-plausible connectivity

## REFERENCES

- [1] E. Objois and N. Vadot, “BIO-465 project source code,” May 2022. [Online]. Available: <https://github.com/ninivert/BIO465-cortical-surround-suppression>

## APPENDIX

### A. Full expression of $\mathbf{h}_f$ in the linear ISN

$$\begin{aligned}
 h_{f0} &= -\frac{I_0 (W_{11}\tau_0 - \tau_0)}{\tau_0 (W_{00}W_{11} - W_{00} - W_{01}W_{10} - W_{11} + 1)} \\
 &\quad + \frac{I_1 W_{01}}{W_{00}W_{11} - W_{00} - W_{01}W_{10} - W_{11} + 1} \\
 h_{f1} &= \frac{I_0 W_{10}}{W_{00}W_{11} - W_{00} - W_{01}W_{10} - W_{11} + 1} \\
 &\quad - \frac{I_1 (W_{00}\tau_1 - \tau_1)}{\tau_1 (W_{00}W_{11} - W_{00} - W_{01}W_{10} - W_{11} + 1)}
 \end{aligned}$$