S1 Appendix: Amplifying the redistribution of somato-dendritic inhibition by the interplay of three interneuron types

Loreen Hertäg\textsuperscript{1,2}, Henning Sprekeler\textsuperscript{1,2},

\textsuperscript{1} Modelling of Cognitive Processes, Berlin Institute of Technology, Berlin, Germany
\textsuperscript{2} Bernstein Center for Computational Neuroscience, Berlin, Germany

* h.sprekeler@tu-berlin.de, loreen.hertaeg@tu-berlin.de

Mathematical analysis of the SOM-VIP motif

In order to characterize the computational repertoire of the SOM-VIP motif, we performed an extensive mathematical analysis of a simplified model that comprises only SOM and VIP cell populations. Moreover, we neglected the rectifying nonlinearity that ensures that firing rates remain positive. We will start with networks that take into account either recurrent connections among both SOM and VIP neurons or adaptation, but ignore short-term plasticity. In these cases, the circuits are described by linear dynamical systems and therefore allow detailed bifurcation analyses. The derivations provide analytical conditions for the parameter boundaries between different computational states of the SOM-VIP motif. We then continue to derive approximations for these boundaries for the nonlinear case that includes short-term plasticity.

Bifurcation analysis: Recurrent inhibition

We consider a network composed of SOM (hereafter only S) and VIP neurons (hereafter only V) that are mutually and fully connected. Furthermore, we assume that each population comprises \( n \) cells. The circuit dynamics can then be described by

\[
\dot{r} = Wr + s
\]

with \( r^T = [r_{S1}, \ldots, r_{Sn}, r_{V1}, \ldots, r_{Vn}] \). \( s \) represents a vector of external stimuli and \( W \) is a block matrix of the form

\[
W = \begin{bmatrix} U_{S\rightarrow S} & U_{S\rightarrow V} \\ U_{V\rightarrow S} & U_{V\rightarrow V} \end{bmatrix}
\]

which contains the connection weights. When recurrence is included, the \((n \times n)\) blocks have the following form

\[
U_{S\rightarrow S} = \begin{bmatrix} -1/\tau_S & -w_S/\tau_S & \ldots & -w_S/\tau_S \\ -w_S/\tau_S & -1/\tau_S & \ldots & -w_S/\tau_S \\ \vdots & \vdots & \ddots & \vdots \\ -w_S/\tau_S & -w_S/\tau_S & \ldots & -1/\tau_S \end{bmatrix}
\]

\[
U_{S\rightarrow V} = \begin{bmatrix} -1/\tau_V & -w_V/\tau_V & \ldots & -w_V/\tau_V \\ -w_V/\tau_V & -1/\tau_V & \ldots & -w_V/\tau_V \\ \vdots & \vdots & \ddots & \vdots \\ -w_V/\tau_V & -w_V/\tau_V & \ldots & -1/\tau_V \end{bmatrix}
\]

\[
U_{V\rightarrow S} = \begin{bmatrix} -w_V/\tau_S \\ -w_v/\tau_S \\ \vdots \\ -w_v/\tau_S \end{bmatrix} J_{n,n}
\]

\[
U_{V\rightarrow V} = \begin{bmatrix} -w_V/\tau_V \\ -w_v/\tau_V \\ \vdots \\ -w_v/\tau_V \end{bmatrix} J_{n,n}
\]
where $J_{n,n}$ is a squared matrix of dimension $n$ with all entries equal one. Note that autapses were not included and that we chose the convention that the weight parameters $w_i$ are positive.

The dynamical properties of the circuit are described by its eigenvalues, which are given by the zero crossings of the characteristic polynomial of the matrix $W$, 

$$0 = \det((U_{S+S} - \lambda \mathbb{1})(U_{V+V} - \lambda \mathbb{1}) - U_{S+V}U_{V+S}),$$  \hspace{1cm} (5)$$

where $\mathbb{1}$ denotes the $n$-dimensional identity matrix. Resolving the determinant yields the following condition for the eigenvalues:

$$0 = \left[\left(\frac{1}{\tau_S} + \lambda\right)\left(\frac{1}{\tau_V} + \lambda\right) + \frac{w_S w_V}{\tau_S \tau_V} - \left(\frac{1}{\tau_S} + \lambda\right) \frac{w_V}{\tau_V} - \left(\frac{1}{\tau_V} + \lambda\right) \frac{w_S}{\tau_S}\right]^{n-1},$$  \hspace{1cm} (6)$$

with the total recurrence strengths $\hat{w}_{mn} = (n-1)w_{mn}$ and the total mutual inhibition strengths $\hat{w}_m = nw_m$. Consequently, the eigenvalues of the linear system are

$$\lambda_{S} = -\frac{1}{\tau_S} + \frac{\hat{w}_S}{(n-1)\tau_S},$$  \hspace{1cm} (7)$$

$$\lambda_{V} = -\frac{1}{\tau_V} + \frac{\hat{w}_V}{(n-1)\tau_V},$$

$$\lambda_{+/-} = -\frac{1}{2} \left(\frac{1 + \hat{w}_S \tau_S}{\tau_S} + \frac{1 + \hat{w}_V \tau_V}{\tau_V}\right) \pm \frac{1}{2} \sqrt{\left[\left(\frac{1}{\tau_V} - \frac{1}{\tau_S}\right) + \left(\frac{\hat{w}_V}{\tau_V} - \frac{\hat{w}_S}{\tau_S}\right)\right]^2 + 4 \frac{\hat{w}_S \hat{w}_V \hat{w}_{VS}}{\tau_S \tau_V \tau_V}},$$  \hspace{1cm} (8)$$

While the first and second eigenvalues $\lambda_{S}, \lambda_{V}$ have an algebraic and geometric multiplicity of $(n-1)$ and describe the dynamics of rate inhomogeneities within the two populations, the eigenvalues $\lambda_{+}$ and $\lambda_{-}$ have an algebraic and geometric multiplicity of one and determine the dynamics of the population rates of the two interneuron types. Because $\hat{w}_{SV}$ and $\hat{w}_{VS}$ are positive, all eigenvalues are real, suggesting the absence of oscillations. For the sake of simplicity, we now analyze a symmetric situation, in which $\tau_S = \tau_V = \tau$, $\hat{w}_S = \hat{w}_V = w_1$, and $\hat{w}_{SV} = \hat{w}_{VS} = w$. By analyzing the sign and the nature (complex or real) of the eigenvalues, we can define five dynamical regimes:

(i) All interneurons can be active and operate in an attenuation regime (no competition, all eigenvalues negative, weak mutual inhibition),

(ii) All interneurons can be active and operate in an amplification regime (no competition, all eigenvalues negative, strong mutual inhibition),

(iii) WTA between SOM and VIP neuron population (either SOM or VIP cells win; $\lambda_{+/-} > 0$, $\lambda_{S/V} < 0$),

(iv) WTA in each population separately (one VIP and one SOM cell survive, $\lambda_{+/-} < 0$, $\lambda_{S/V} > 0$ positive),

(v) Total WTA (only one single neuron is active, all eigenvalues positive).

The transition between the attenuation and the amplification regime ((i) and (ii)) is determined by the condition that the amplification index (equation (24) in main text) is
equal to one (for the symmetric case \( \dot{w}_{VV} = \dot{w}_{SS} = \dot{w}_r \) and \( \dot{w}_{SV} = \dot{w}_{VS} = \dot{w} \)). The transition to a WTA regime within each neuron population emerges when \( \lambda_S = \lambda_V = \lambda \geq 0 \), yielding the condition

\[
\dot{w}_r \geq (n - 1).
\]

Moreover, the transition to a WTA regime between SOM and VIP neurons occurs when \( \lambda_+ \geq 0 \), which yields

\[
\dot{w} \geq 1 + \dot{w}_r.
\]

**Bifurcation analysis: Adaptation**

We extend the rate dynamics (see equation (1)) with linear differential equations describing the evolution of an adaptation current in order to derive qualitative changes in the bifurcation structure when adaptation (instead of recurrent inhibition) is present. The \((4 \times n)\)-dimensional state vector \( \mathbf{r} \) now includes the adaptation variables and is given by

\[
\mathbf{r}^T = [r_S, a_S, \ldots, r_S, a_S, r_V, a_V, \ldots, r_V, a_V].
\]

Again, the dynamical system can be written in the format of equations (1) and (2), but the matrices \( U_{m=n} \) \((m, n \in \{S, V\})\) now themselves become block matrices:

\[
U_{S+S} = \begin{bmatrix}
M_0 & 0 & \cdots & 0 \\
0 & M_0 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & M_0
\end{bmatrix},
U_{S+V} = \begin{bmatrix}
N_0 & N_0 & \cdots & N_0 \\
N_0 & N_0 & \cdots & N_0 \\
\vdots & \vdots & \ddots & \vdots \\
N_0 & N_0 & \cdots & N_0
\end{bmatrix},
U_{V+S} = \begin{bmatrix}
P_0 & P_0 & \cdots & P_0 \\
P_0 & P_0 & \cdots & P_0 \\
\vdots & \vdots & \ddots & \vdots \\
P_0 & P_0 & \cdots & P_0
\end{bmatrix},
U_{V+V} = \begin{bmatrix}
Q_0 & 0 & \cdots & 0 \\
0 & Q_0 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & Q_0
\end{bmatrix}.
\]

Here, \( M_0, N_0, P_0 \) and \( Q_0 \) are \((2 \times 2)\) matrices given by

\[
M_0 = \begin{bmatrix}
-1/\tau_S & -1/\tau_S \\
bs/\tau_{a,S} & -1/\tau_{a,S}
\end{bmatrix},
N_0 = \begin{bmatrix}
-w_{SV}/\tau_S & 0 \\
0 & 0
\end{bmatrix},
Q_0 = \begin{bmatrix}
-1/\tau_V & -1/\tau_V \\
-bv/\tau_{a,V} & -1/\tau_{a,V}
\end{bmatrix},
P_0 = \begin{bmatrix}
-w_{SV}/\tau_V & 0 \\
0 & 0
\end{bmatrix}
\]

with the adaptation strengths \( b_S \) and \( b_V \), as well as adaptation time constants \( \tau_{a,S} \) and \( \tau_{a,V} \). As the block matrices do not commute, the eigenvalue condition for the characteristic polynomial is now given by

\[
0 = \det(W - \lambda I) = \det(U_{S+S} - \lambda I) \\
\cdot \det([U_{V+V} - \lambda I] - U_{V+S} [U_{S+S} - \lambda I]^{-1} U_{S+V}).
\]

After some straightforward linear algebra, the eigenvalues are given by the zeros of the following equation:

\[
0 = \left[ \left( \frac{1}{\tau_S} + \lambda \right) \left( \frac{1}{\tau_{a,S}} + \lambda \right) + \frac{b_S}{\tau_{S,a,S}} \right]^{(n-1)} \left[ \left( \frac{1}{\tau_V} + \lambda \right) \left( \frac{1}{\tau_{a,V}} + \lambda \right) + \frac{b_V}{\tau_{V,a,V}} \right]^{(n-1)} \\
\cdot \left[ \left( \frac{1}{\tau_S} + \lambda \right) \left( \frac{1}{\tau_{a,S}} + \lambda \right) + \frac{b_S}{\tau_{S,a,S}} \right] \left\{ \left( \frac{1}{\tau_V} + \lambda \right) \left( \frac{1}{\tau_{a,V}} + \lambda \right) + \frac{b_V}{\tau_{V,a,V}} \right\} \\
- \frac{\dot{w}_{SV}\dot{w}_{VS}}{\tau_S\tau_V} \left( \frac{1}{\tau_{a,S}} + \lambda \right) \left( \frac{1}{\tau_{a,V}} + \lambda \right).
\]
For the sake of simplicity and comparability, we again consider the symmetric case 
\( \tau_S = \tau_V = \tau, \ w_{SV} = w_{VS} = w, \ b_S = b_V = b \) and \( \tau_{a,S} = \tau_{a,V} = \tau_a \). This simplifies the derivation of the eigenvalues considerably. We obtain three pairs of eigenvalues 

\[
\begin{align*}
\lambda_{1,\pm} &= \frac{1}{2} \left( -\left( \frac{1}{\tau} + \frac{1}{\tau_a} \right) \pm \sqrt{\left( \frac{1}{\tau} - \frac{1}{\tau_a} \right)^2 - \frac{4b}{\tau \tau_a}} \right), \\
\lambda_{2,\pm} &= \frac{1}{2} \left( -\left( \frac{1}{\tau} + \frac{1}{\tau_a} \mp \frac{\hat{w}}{\tau} \right) \pm \sqrt{\left( \frac{1}{\tau} - \frac{1}{\tau_a} \mp \frac{\hat{w}}{\tau} \right)^2 - \frac{4b}{\tau \tau_a}} \right), \\
\lambda_{3,\pm} &= \frac{1}{2} \left( -\left( \frac{1}{\tau} + \frac{1}{\tau_a} + \frac{\hat{w}}{\tau} \right) \pm \sqrt{\left( \frac{1}{\tau} - \frac{1}{\tau_a} + \frac{\hat{w}}{\tau} \right)^2 - \frac{4b}{\tau \tau_a}} \right). 
\end{align*}
\]

The first two eigenvalues \( \lambda_{1,\pm} \) and \( \lambda_{1,-} \) have both algebraic and geometric multiplicity of \((2n - 2)\) and define the rate and adaptation dynamics of rate inhomogeneities within the two interneuron populations. Both eigenvalues are strictly negative for all parameters possible. Hence, in contrast to recurrence, adaptation does not lead to pathological states in which a single cell within one population silences the others (see conditions (iv) and (v) in section above). The last four eigenvalues have algebraic and geometric multiplicity of one and describe the interaction between the two populations. It can be easily seen that \( \text{Re}(\lambda_{3,\pm}) < 0 \), so that instabilities – like WTA regimes – can only arise from the two eigenvalues \( \lambda_{2,\pm} \). Depending on the sign and nature of these eigenvalues, we find four dynamical regimes:

(i) All interneurons can be active and operate in an attenuation regime (all eigenvalues real and negative, weak mutual inhibition),

(ii) All interneurons can be active and operate in an amplification regime (all eigenvalues real and negative, strong mutual inhibition),

(iii) WTA between SOM and VIP neuron population, described by the condition

\[
b \leq \hat{w} - 1, \tag{19}\]

(iv) An oscillatory WTA regime (osc WTA) in which SOM and VIP neurons alternate between active and inactive states. It depends on two conditions

\[
\begin{align*}
b &> \hat{w} - 1, \\
\hat{w} &> 1 + \frac{\tau}{\tau_a}.
\end{align*} \tag{20}
\]

The derivation of the amplification index \( A \) for a network with adaptation is analogous to the derivation with recurrent inhibition, and merely requires to replace the total recurrence strength \( \hat{w} \), by the strength \( b \) of adaptation. Hence, the transition between the attenuation and the amplification regime ((i) and (ii)) is determined by this minor modification of equation \( \text{(24)} \) (see main text).

Strictly speaking, the last regime (iv) has to be separated into two subregimes, in which the two eigenvalues \( \lambda_{2,\pm} \) either (a) form a complex conjugate pair with positive real part or (b) are both real and positive. In the former regime, an approximation of the oscillation frequency \( f \) can be obtained from the imaginary part of the eigenvalues:

\[
f = \frac{1}{4\pi} \sqrt{\frac{4b}{\tau \tau_a}} - \left( \frac{1}{\tau} - \frac{1}{\tau_a} - \frac{\hat{w}}{\tau} \right)^2. \tag{21}\]
The oscillation frequency follows a square root function in \( b \) with a scaling factor determined by \( \tau_a \), and an offset controlled by both \( \hat{w} \) and \( \tau_a \). The parameter regime (a), in which an approximation for the oscillation frequency can be obtained from a linear analysis is limited by the condition

\[
b > \frac{\tau_a}{4} \left( \frac{1}{\tau} - \frac{1}{\tau_a} - \frac{\hat{w}}{\tau} \right)^2.
\] (22)

In the second regime (b), the circuit nevertheless displays oscillations when simulated. These rely on the rectifying nonlinearity for the firing rates, so their frequency cannot be calculated within the present linear analysis.

**Bifurcation analysis: Adaptation and short-term facilitation**

When synaptic adaptation (STF) is included in the model, the dynamical system equations become nonlinear and mathematically intractable. Thus, we can only derive approximate expressions for the transitions between the computational regimes (see (i) to (iv) in section above).

In zero-order approximation, we assume that the facilitation variable \( u_{ij} \) can be replaced by its steady-state value \( u_{\infty} \), for which we derive an expression below. The boundaries given by equations (19) and (20) can then be replaced by

\[
b = u_{\infty} \hat{w} - 1, \quad (23)
\]

\[
u_{\infty} \hat{w} = 1 + \frac{\tau}{\tau_a}. \quad (24)
\]

These approximations allow to deduce the qualitative changes in the transition boundaries. As \( u_{\infty} \geq 1 \), the slope of equation (23) increases, leading to an enlarged WTA regime (i.e., switch-like state). Moreover, the transition to an oscillatory switch is pushed towards smaller mutual inhibition strengths and requires stronger adaptation (cf. equations (23)-(24)).

To qualify the regime boundaries, we have to derive an expression for the steady-state value \( u_{\infty} \) as a function of the parameters \( b \) and \( \hat{w} \). To this end, we identify a self-consistent solution of the equations for \( u_{\infty} \) and the steady state of the population rates of the two interneuron types.

The steady state of the population rates \( r_{V/S} := \frac{1}{N_{V/S}} \sum_i r_{V/S,i} \) is given by the conditions

\[
r_V = \bar{s} - \hat{w} u_{\infty}^V r_S - b r_V, \quad (25)
\]

\[
r_S = \bar{s} - \hat{w} u_{\infty}^S r_V - b r_S. \quad (26)
\]

Here, \( \bar{s} \) denotes the averaged input. \( u_{\infty}^V \) and \( u_{\infty}^S \) represent the steady-state facilitation variable for SOM\(\rightarrow\)VIP and VIP\(\rightarrow\)SOM, respectively, which are given by

\[
u_{\infty}^{V/S} = \frac{1 + \tau \tau r_{V/S}}{1 + U_s \tau r_{V/S}}. \quad (27)
\]

To express \( u_{\infty} \) in terms of \( b \) and \( \hat{w} \), we have to resolve equation (26) for the population rates and insert these expressions into equation (27). Because the resulting equations are lengthy and non-informative, we consider two approximations. In the non-WTA regime, as intrinsic and synaptic parameters as well as the external stimulation are taken to be equal, we assume that the population firing rates of SOM and VIP neurons
are approximately equal on average \( r_V \approx r_S \equiv r \). This leads to the following expression for the population rates \( r \) in terms of the system parameters:

\[
\tau_t U_s \bar{s} - 1 - \tau_t \bar{s} + \sqrt{\tau_t U_s \bar{s} - 1 - \tau_t \bar{s}}^2 + 4\bar{s} \tau f \tau_t U_s \bar{s} \]

\[
\frac{1}{2\tau_t (1 + b)}
\]

By inserting equations (27) and (28) into equation (24), we obtained a good approximation of the transition boundary between the amplification regimes and the oscillatory switch state.

For the transition to a WTA regime, the assumption of equal rates is violated. To derive an expression for this transition boundary, we therefore replace the assumption of equal rates by an assumption of a WTA state, in which one of the rates is zero. Because all parameters are assumed to be symmetric, we can assume without loss of generality that the VIP population is the winner. We can then obtain conditions for the boundary by setting \( r_S = 0 \) in equations (25) and (26) and inserting equation (27) into equation (26):

\[
\tau_t U_s \bar{s} - \bar{w} + \sqrt{\tau_t U_s \bar{s} - \bar{w}}^2 + 4\bar{w} \tau_t \bar{s}
\]

\[
\frac{1}{2\tau_t (1 + b)}
\]

Solving this equation for the adaptation strength \( b \), we obtain

\[
b = \frac{2\bar{w} \tau_t \bar{s}}{\tau_t U_s \bar{s} - \bar{w} + \sqrt{\tau_t U_s \bar{s} - \bar{w}}^2 + 4\bar{w} \tau_t \bar{s}} - 1.
\]

This approximation predicts the transition to the WTA regime qualitatively and quantitatively very well.

For deriving the transition boundary between attenuation and amplification, we note that the amplification threshold \( A = 1 \) is satisfied when the slopes for the full network and the reference network are equal,

\[
\frac{dr_{S}^{\text{full}}}{dx_{\text{mod}}} = \frac{dr_{S}^{\text{ref}}}{dx_{\text{mod}}}.
\]

The steady state equation for the population rates of the SOM neurons in the reference and full network, respectively, are given by

\[
(1 + b) r_S^{\text{ref}} = \bar{s} - x_{\text{mod}},
\]

\[
(1 + b) r_S^{\text{full}} = \bar{s} - u_{\infty}^V r_V,
\]

where \( u_{\infty}^V \) is given in equation (27). The population rate \( r_V \) of the VIP neurons in the full network obeys

\[
(1 + b) r_V = \bar{s} + x_{\text{mod}} - u_{\infty}^S r_S^{\text{full}}.
\]

Taking the derivative of the steady-state equations (32) and (33) for full and reference network with respect to \( x_{\text{mod}} \) yields

\[
\frac{dr_{S}^{\text{ref}}}{dx_{\text{mod}}} = -\frac{1}{1 + b},
\]

\[
\frac{dr_{S}^{\text{full}}}{dx_{\text{mod}}} = -\frac{w}{1 + b} \frac{(1 + 2\tau_t r_V)(1 + U_s \tau_t r_V) - (1 + \tau_t r_V)U_s \tau_t r_V}{(1 + U_s \tau_t r_V)^2} \frac{dr_V}{dx_{\text{mod}}}.
\]
By taking the derivative of equation (34), we can express \( \frac{dr_V}{dx_{\text{mod}}} \) as a function of \( \frac{d_{r_S}^{\text{full}}}{dx_{\text{mod}}} \) and \( i_S^{\text{full}} \). Furthermore, we assume again that \( r_V \approx r_S \equiv r \) (cf. equation (28)). Solving for \( \frac{dr_S^{\text{full}}}{dx_{\text{mod}}} \), the condition (31) finally yields

\[
(1 + b)^2 - \hat{w}^2 \gamma^2 = \hat{w} \gamma (1 + b),
\]

with

\[
\gamma = \frac{(1 + 2\tau r)(1 + U_s \tau r) - (1 + \tau r)U_s \tau r}{(1 + U_s \tau r)^2}.
\]

The combination of equations (37), (38) and (28) provide an accurate prediction of the amplification threshold when STF and adaptation are present.

References

1. Silvester, J. R. Determinants of block matrices. *The Mathematical Gazette* **84**, 460–467 (2000).

2. Murayama, M. *et al.* Dendritic encoding of sensory stimuli controlled by deep cortical interneurons. *Nature* **457**, 1137 (2009).