



university of
 groningen

faculty of science
 and engineering

Including STDP to eligibility propagation in multi-layer recurrent spiking neural networks.

Werner van der Veen
(w.k.van.der.veen.2@student.rug.nl)

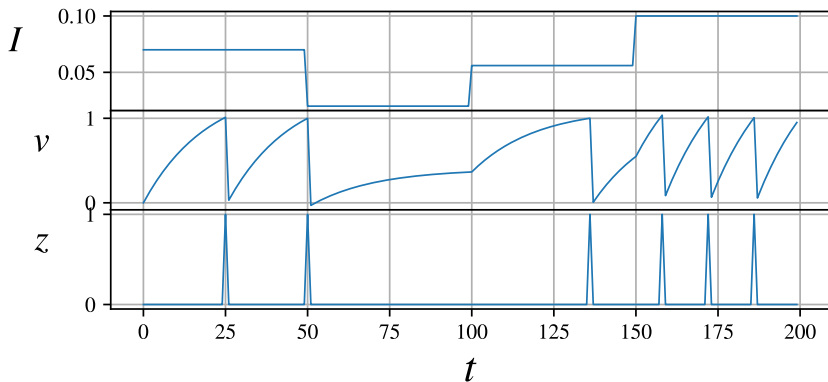
April 29, 2021

Introduction – context

- Biological inspiration to AI
- ANNs and backpropagation
- Deep learning
 - high energy consumption
- Fundamental differences between DL and biological learning
 - Continuous vs. binary communication
 - Backpropagation/BPTT vs. learning signals

Introduction – SNNs

- Binary spikes \rightarrow no backpropagation
- Good performance, but not competitive with DL
- Leaky integrate-and-fire (LIF) neuron



Introduction – neuromorphic computing

- Physical embedding of neural network in an analog medium
- Colocalized memory and computation
 - no backpropagation
- SNNs as ideal biologically inspired NC paradigm for efficient computation
- Analog SNNs are extremely fast and energy efficient
- Problem: learning rule must be *local* and *online*
 - revalue biological learning

Introduction – biological learning

- Hebbian learning
 - runaway excitation
- STDP
 - how to teach?
- Learning signals: three-factor Hebbian learning, R-STDP
- Biological learning signals: neurotransmitters
 - credit assignment
- Eligibility traces

Introduction – eligibility propagation

- Mathematical approximation to BPTT
- Local learning signals and eligibility traces
- Applicable to any SNN topology and multiple neuron models
- Also applicable to neuromorphic VLSIs
- Competitive with LSTMs on phone classification
- E-prop: only ALIF so far
- New research: STDP-LIF and Izhikevich neuron display STDP

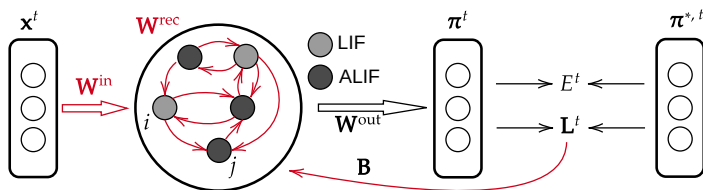
Introduction – this research

Research question:

Does including STDP-like behavior in e-prop lead to faster and more accurate phone classification?

- Reproduce results of original e-prop paper
- Extend STDP-LIF to STDP-ALIF
- Experimentally verify the STDP-ALIF and Izhikevich neurons on phone classification task

Methods – technical framework (e-prop)

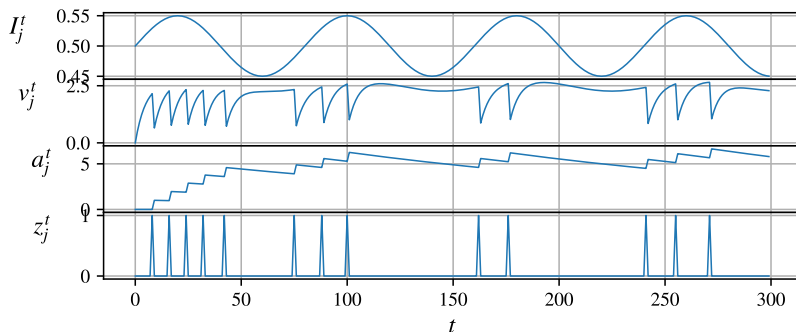


E-prop model $\mathcal{M} = \langle M, f \rangle$

$$\mathbf{h}_j^t = M(\mathbf{h}_j^{t-1}, \mathbf{z}^{t-1}, \mathbf{x}^t, \mathbf{W}_j) \quad (1)$$

$$\mathbf{z}_j^t = f(\mathbf{h}_j^t) \quad (2)$$

Methods – the ALIF neuron model



$$z_j^t = H(v_j^t - v_{\text{th}} - \beta a_j^t) \quad (3)$$

$$v_j^{t+1} = \alpha v_j^t + \sum_{i \neq j} W_{ji}^{\text{rec}} z_i^t + \sum_i W_{ji}^{\text{in}} x_i^{t+1} - z_j^t v_{\text{th}} \quad (4)$$

$$a_j^{t+1} = \rho a_j^t + z_j^t \quad (5)$$

Methods – e-prop

$$\frac{dE}{dW_{ji}} = \sum_t L_j^t e_{ji}^t$$

$$\frac{dE}{dW_{ji}} = \sum_t \frac{dE}{dz_j^t} \frac{\partial z_j^t}{\partial \mathbf{h}_j^t} \underbrace{\sum_{t \geq t'} \underbrace{\frac{\partial \mathbf{h}_j^t}{\partial \mathbf{h}_j^{t-1}} \cdots \frac{\partial \mathbf{h}_j^{t+1}}{\partial \mathbf{h}_j^{t'}}}_{\epsilon_{ji}^t}}_{e_{ji}^t} \cdot \frac{\partial \mathbf{h}_j^{t'}}{\partial W_{ji}}$$

Methods – e-prop for ALIF

Hidden state:

$$\mathbf{h}_j^t = \begin{pmatrix} v_j^t \\ a_j^t \end{pmatrix} \quad (6)$$

Pseudo-derivative:

$$\psi_j^t = \gamma \max \left(0, 1 - \left| \frac{v_j^t - v_{\text{th}} - \beta a_j^t}{v_{\text{th}}} \right| \right) \quad (7)$$

Eligibility vector:

$$\begin{pmatrix} \epsilon_{ji,v}^{t+1} \\ \epsilon_{ji,a}^{t+1} \end{pmatrix} = \begin{pmatrix} \alpha \cdot \epsilon_{ji,v}^t + z_i^{t-1} \\ \psi_j^t \epsilon_{ji,v}^t + (\rho - \psi_j^t \beta) \epsilon_{ji,a}^t \end{pmatrix} \quad (8)$$

Eligibility trace:

$$e_{ji}^t = \psi_j^t (\epsilon_{ji,v}^t - \beta \epsilon_{ji,a}^t) \quad (9)$$

Methods – e-prop weight update

$$\Delta W_{ji} = -\eta \underbrace{\sum_t \sum_k B_{jk} (\pi_k^t - \pi_k^{*,t})}_{=L_j^t} \underbrace{\sum_{t' \leq t} \kappa^{t'-t} e_{ji}^{t'}}_{\stackrel{\text{def}}{=} \bar{e}_{ji}^t} \quad (10)$$

where $\hat{y}_k^t = \kappa \hat{y}_k^{t-1} + \sum_j W_{kj}^{\text{out}} z_j^t$,

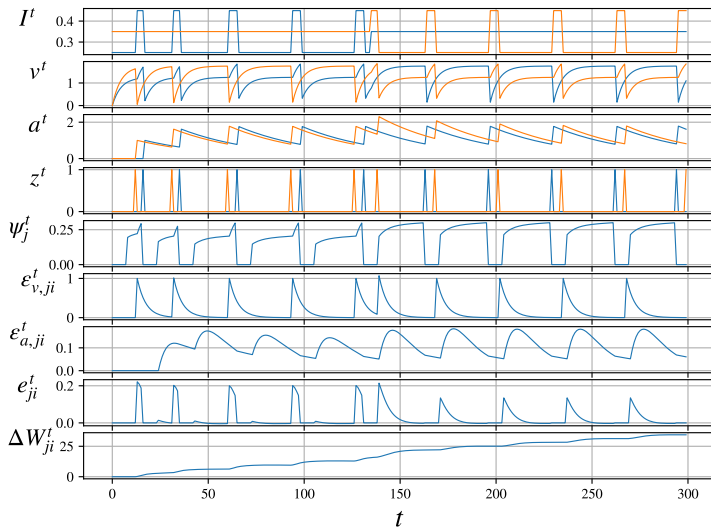
$y_k^t = \hat{y}_k^t + b_k$

and $\pi_k^t = \sigma_k(y_1^t, \dots, y_K^t)$.

$$\Delta W_{kj}^{\text{out}} = -\eta \sum_t (\pi_k^t - \pi_k^{*,t}) \sum_{t' \leq t} \kappa^{t'-t} z_j^t \quad (11)$$

$$\Delta b_k = -\eta \sum_t (\pi_k^t - \pi_k^{*,t}). \quad (12)$$

Methods – Visualizing the ALIF neuron



Methods – STDP-ALIF

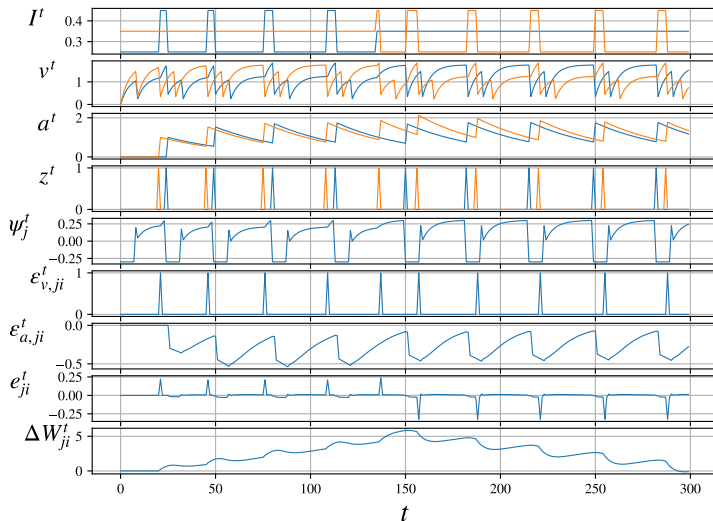
- New neuron model:

$$v_j^{t+1} = \alpha v_j^t + \sum_{i \neq j} W_{ji}^{\text{rec}} z_i^t + \sum_i W_{ji}^{\text{in}} I - z_j^t \alpha v_j^t - z_j^{t-\delta t_{\text{ref}}} \alpha v_j^t \quad (13)$$

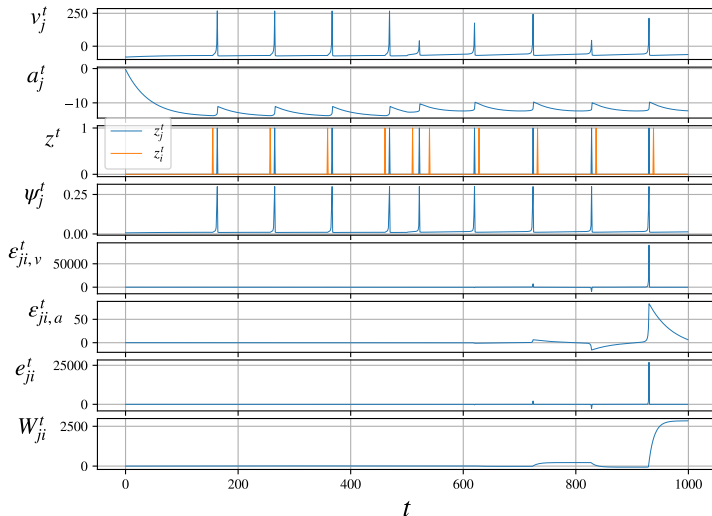
$$\psi_j^t = \begin{cases} -\gamma & \text{if } t - t_{z_j} < \delta t_{\text{ref}} \\ \gamma \max\left(0, 1 - \left|\frac{v_j^t - v_{\text{th}}}{v_{\text{th}}}\right|\right) & \text{otherwise} \end{cases} \quad (14)$$

- Decreases v_j^t after refractory period, and clamps ψ_j^t to $-\gamma$ instead of to 0.

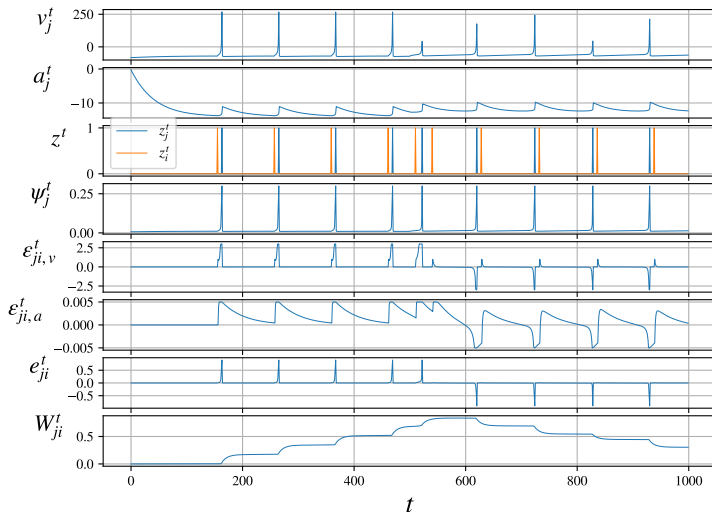
Methods – STDP-ALIF



Methods – Izhikevich

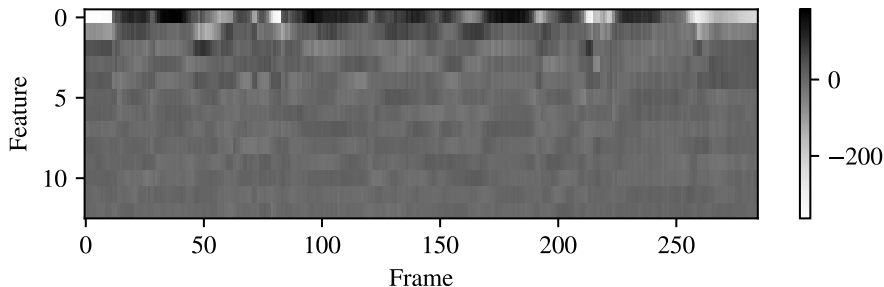


Methods – Izhikevich fixed



Methods – Task

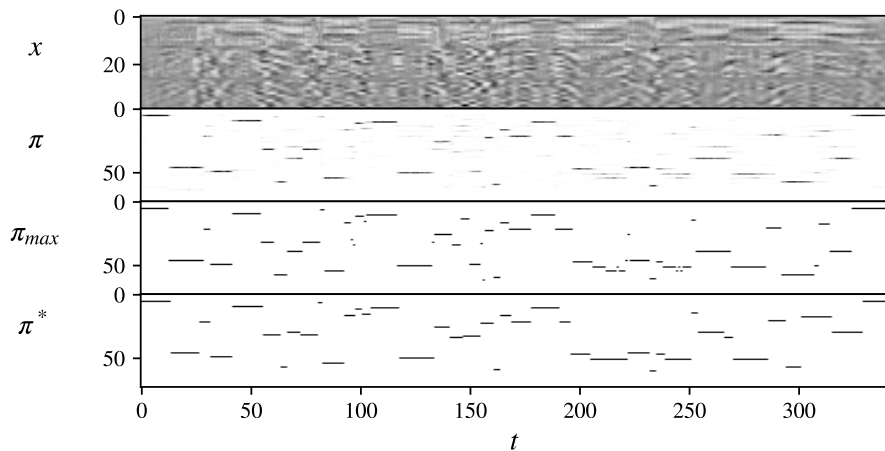
- Classify frame-wise phones from speech signals.
- There are 61 different class labels, and approximately 4000 speech sentences containing around 200-300 frames each.
- Speech is preprocessed into MFCCs (2–13) with their 1st and 2nd deltas, standardized per channel



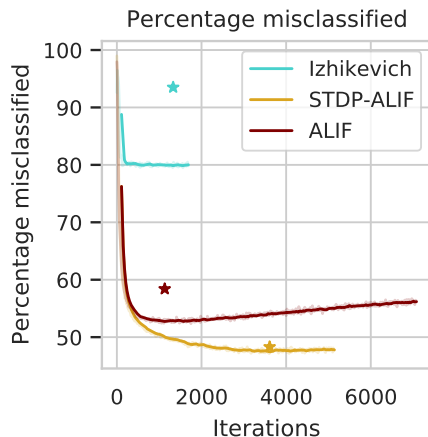
Methods – Other settings

- Adam optimizer for e-prop
- Firing rate regularization
- L2 regularization

Results – Example

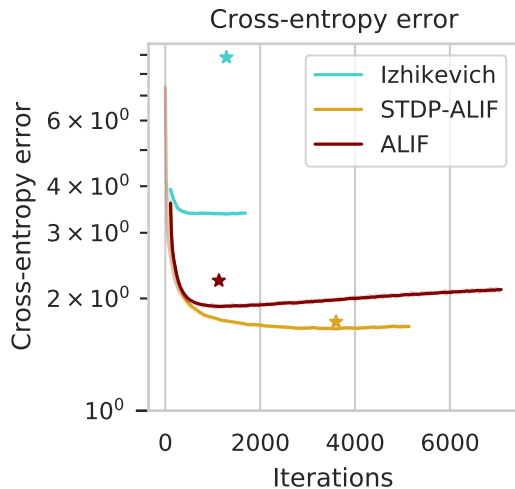


Results – Accuracy per neuron type



Test scores: ALIF = 58.4%, STDP-ALIF = 48.3%, Izhikevich = 93.5%

Results – Cross-entropy per neuron type



Discussion

- Including STDP in e-prop improves classification performance and efficiency (but not always!).
 - one of the best biologically plausible learning algorithms for RSNNs.
- Good trade-off between running cost and performance.
- More bioplausibility
 - easier in neuromorphic hardware but not always better performance.
- More research is needed on applying e-prop in other types of learning tasks, and many ideas exist that may improve performance even more.

Discussion – future research

- Smoothen output
- Distributed parameters
- Custom connectivity graphs
- Dynamic pruning and growing
- Synaptic delay

Thank you for listening!

$$e_{ji}^t = \epsilon_{ji}^t \cdot \frac{\partial z_j^t}{\partial \mathbf{h}_j^t} \quad (15)$$

$$\begin{pmatrix} \epsilon_{ji,v}^{t+1} \\ \epsilon_{ji,a}^{t+1} \end{pmatrix} = \begin{pmatrix} \frac{\partial v_j^{t+1}}{\partial v_j^t} & \frac{\partial v_j^{t+1}}{\partial a_j^t} \\ \frac{\partial a_j^{t+1}}{\partial v_j^t} & \frac{\partial a_j^{t+1}}{\partial a_j^t} \end{pmatrix} \cdot \begin{pmatrix} \epsilon_{ji,v}^t \\ \epsilon_{ji,a}^t \end{pmatrix} + \begin{pmatrix} \frac{\partial v_j^{t+1}}{\partial W_{ji}^t} \\ \frac{\partial a_j^{t+1}}{\partial W_{ji}^t} \end{pmatrix} \quad (16)$$

$$\pi_k^t = \sigma_k(y_1^t, \dots, y_K^t) = \frac{\exp(y_k^t)}{\sum_{k'} \exp(y_{k'}^t)} \quad (17)$$

$$E = - \sum_{t,k} \pi_k^{*,t} \log \pi_k^t \quad (18)$$

$$L_j^t = \frac{\partial E}{\partial z_j^t} = \sum_k B_{jk} \sum_{t' \geq t} \left(\pi_k^{t'} - \pi_k^{*,t'} \right) \kappa^{t'-t} \quad (19)$$

Initialize weights from $\mathcal{N}(0, \sqrt{N_{\text{afferent}}})$

$$\frac{dE}{dW_{ji}} = \sum_{t' \leq T} \frac{dE}{d\mathbf{h}_j^{t'}} \cdot \frac{\partial \mathbf{h}_j^{t'}}{\partial W_{ji}} = \sum_t \frac{dE}{dz_j^t} \cdot \left[\frac{dz_j^t}{dW_{ji}} \right]_{\text{local}} \quad (20)$$

$$\frac{dE}{d\mathbf{h}_j^{t'}} = \underbrace{\frac{dE}{dz_j^{t'}} \frac{\partial z_j^{t'}}{\partial \mathbf{h}_j^{t'}}}_{L_j^{t'}} + \frac{dE}{d\mathbf{h}_j^{t'+1}} \frac{\partial \mathbf{h}_j^{t'+1}}{\partial \mathbf{h}_j^{t'}} \quad (21)$$

$$\frac{dE}{dW_{ji}} = \sum_{t'} \left(L_j^{t'} \frac{\partial z_j^{t'}}{\partial \mathbf{h}_j^{t'}} + \left(L_j^{t'+1} \frac{\partial z_j^{t'+1}}{\partial \mathbf{h}_j^{t'+1}} + (\dots) \frac{\partial \mathbf{h}_j^{t'+2}}{\partial \mathbf{h}_j^{t'+1}} \right) \frac{\partial \mathbf{h}_j^{t'+1}}{\partial \mathbf{h}_j^{t'}} \right) \cdot \frac{\partial \mathbf{h}_j^{t'}}{\partial W_{ji}} \quad (22)$$

$$\frac{dE}{dW_{ji}} = \sum_{t'} \sum_{t \geq t'} L_j^t \frac{\partial z_j^t}{\partial \mathbf{h}_j^t} \frac{\partial \mathbf{h}_j^t}{\partial \mathbf{h}_j^{t-1}} \dots \frac{\partial \mathbf{h}_j^{t+1}}{\partial \mathbf{h}_j^{t'}} \cdot \frac{\partial \mathbf{h}_j^{t'}}{\partial W_{ji}} \quad (23)$$

$$\frac{dE}{dW_{ji}} = \sum_t \frac{dE}{dz_j^t} \frac{\partial z_j^t}{\partial \mathbf{h}_j^t} \underbrace{\sum_{t \geq t'} \frac{\partial \mathbf{h}_j^t}{\partial \mathbf{h}_j^{t-1}} \dots \frac{\partial \mathbf{h}_j^{t+1}}{\partial \mathbf{h}_j^{t'}} \cdot \frac{\partial \mathbf{h}_j^{t'}}{\partial W_{ji}}}_{\epsilon_{ji}^t} \quad (24)$$

$\underbrace{\hspace{10em}}_{e_{ji}^t}$

$$v_j^{t+1} = \alpha v_j^t + \sum_{i \neq j} W_{ji}^{\text{rec}} z_i^t + \sum_i W_{ji}^{\text{in}} x_i^{t+1} - z_j^t v_{\text{th}} \quad (25)$$

$$a_j^{t+1} = \rho a_j^t + z_j^t \quad (26)$$

$$z_j^t = H(v_j^t - v_{\text{th}} - \beta a_j^t) \quad (27)$$

$$\frac{\partial v_j^{t+1}}{\partial v_j^t} = \alpha \quad (28)$$

$$\frac{\partial v_j^{t+1}}{\partial a_j^t} = 0 \quad (29)$$

$$\frac{\partial a_j^{t+1}}{\partial v_j^t} = \psi_j^t \quad (30)$$

$$\frac{\partial a_j^{t+1}}{\partial a_j^t} = \rho - \psi_j^t \beta \quad (31)$$

$$\frac{\partial z_j^t}{\partial \mathbf{h}_j^t} = \begin{pmatrix} \frac{\partial z_j^t}{\partial v_j^t} \\ \frac{\partial z_j^t}{\partial a_j^t} \end{pmatrix} = \begin{pmatrix} \psi_j^t \\ -\beta \psi_j^t \end{pmatrix} \quad (32)$$

STDP-ALIF:

$$\frac{\partial v_{rj}^{t+1}}{\partial v_{rj}^t} = \alpha - z_{rj}^t \alpha - \alpha z_{rj}^{t-\delta t_{\text{ref}}} \quad (33)$$

$$= \alpha \left(1 - z_{rj}^t - z_{rj}^{t-\delta t_{\text{ref}}} \right) \quad (34)$$

$$\epsilon_{rji}^{t+1} = \frac{\partial \mathbf{h}_{rj}^{t+1}}{\partial \mathbf{h}_{rj}^t} \cdot \epsilon_{rji}^t + \frac{\partial \mathbf{h}_{rj}^{t+1}}{\partial W_{rji}} \quad (35)$$

$$= \begin{pmatrix} \alpha \left(1 - z_{rj}^t - z_{rj}^{t-\delta t_{\text{ref}}} \right) \epsilon_{rji,v}^t + z_{ri}^{t-1} \\ \psi_{rj}^t \epsilon_{rji,v}^t + (\rho - \psi_{rj}^t \beta) \epsilon_{rji,a}^t \end{pmatrix} \quad (36)$$

Izhikevich:

$$v' = 0.04v^2 + 5v + 140 - a + I \quad (37)$$

$$a' = 0.004v - 0.02a \quad (38)$$

$$\tilde{v}_{rj}^t = v_{rj}^t - (v_{rj}^t + 65) z_{rj}^t \quad (39)$$

$$\tilde{a}_{rj}^t = a_{rj}^t + 2z_{rj}^t \quad (40)$$

$$v_{rj}^{t+1} = \tilde{v}_{rj}^t + 0.04 (\tilde{v}_{rj}^t)^2 + 5\tilde{v}_{rj}^t + 140 - \tilde{a}_{rj}^t + I_{rj}^t \quad (41)$$

$$a_{rj}^{t+1} = \tilde{a}_{rj}^t + 0.004\tilde{v}_{rj}^t - 0.02\tilde{a}_{rj}^t \quad (42)$$

$$\epsilon_{rji}^{t+1} = \begin{pmatrix} (1 - z_{rj}^t) (6 + 0.08v_{rj}^t) \epsilon_{rji,v}^t - \epsilon_{rji,a}^t + z_{ri}^{t-1} \\ 0.004 (1 - z_{rj}^t) \epsilon_{rji,v}^t + 0.98\epsilon_{rji,a}^t \end{pmatrix} \quad (43)$$

$$e_{rji}^t = \psi_{rj}^t \epsilon_{rji,v}^t \quad (44)$$