An energy-based model of folded autoencoders for unsupervised learning in cortical hierarchies

D. Dold[1,2], J. Sacramento[3], A. F. Kungl[1,2], W. Senn[2], M. A. Petrovici[1,2]

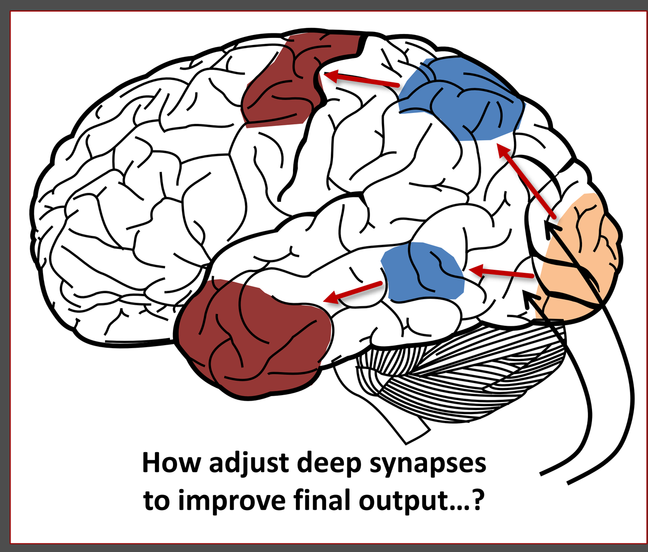
[1]Heidelberg University, Kirchhoff-Institute for Physics, [2]University of Bern, Department of Physiology, [3]Institute of Neuroinformatics, UZH / ETH Zurich.

The Manfred Stärk Foundation

Motivation

Whether the brain uses an optimization scheme like backprop to guide synaptic plasticity in deep hierarchical cortical areas is still an open question.

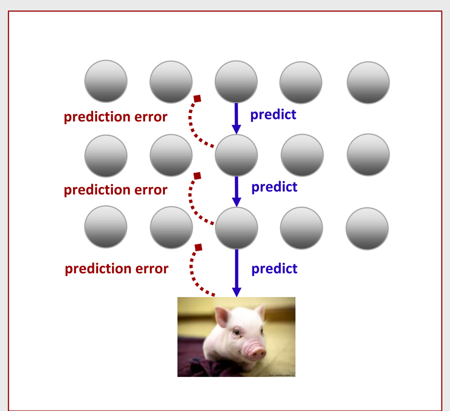
Recently, **several models explaining how backprop might be realized in the cortex have been proposed**, using predictive coding[1], inhibitory microcircuits[2,3] as well as energy-based[1,4] and Lagrangian neurodynamics[5].



Here, **we extend these models to unsupervised learning** and bidirectional (supervised and unsupervised) learning **while maintaining a high degree of biological plausibility**.

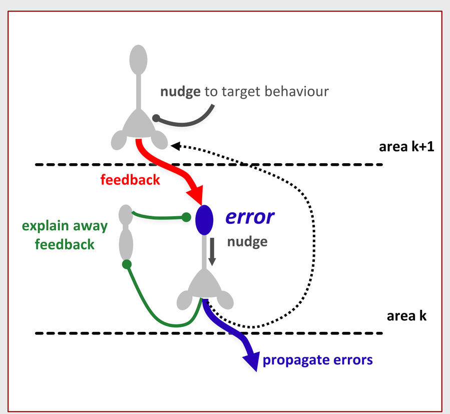
1. Models of error backpropagation

1. Predictive coding[1]



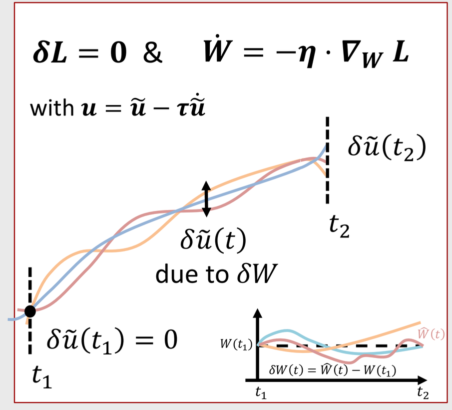
- invert architecture and add input → **supervised learning, approximately backprop**
- **needs phases** for learning, plasticity is only active when the network is stationary
- **high-level**, no implementation details

2. Dendritic microcircuit[3]

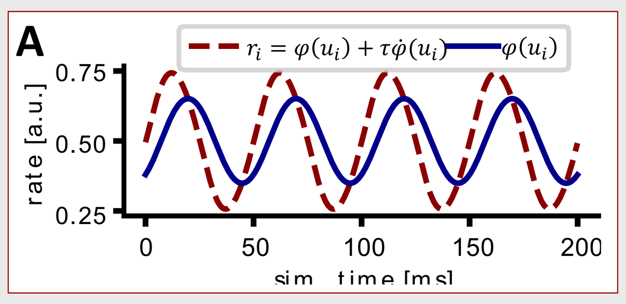


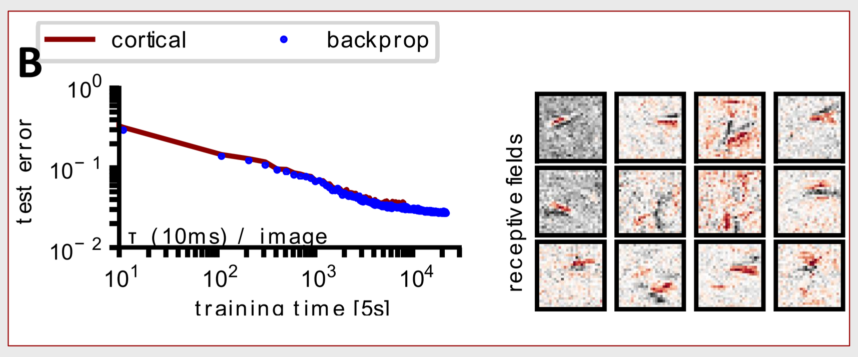
- **apical compartments encode prediction error**
- errors calculated via inhibitory microcircuit
- microcircuit weights trainable to cancel top-layer feedback, **no weight transport**
- neurons hold both forward and error information
- **requires phases** for learning

3. Neuronal Least Action[5]

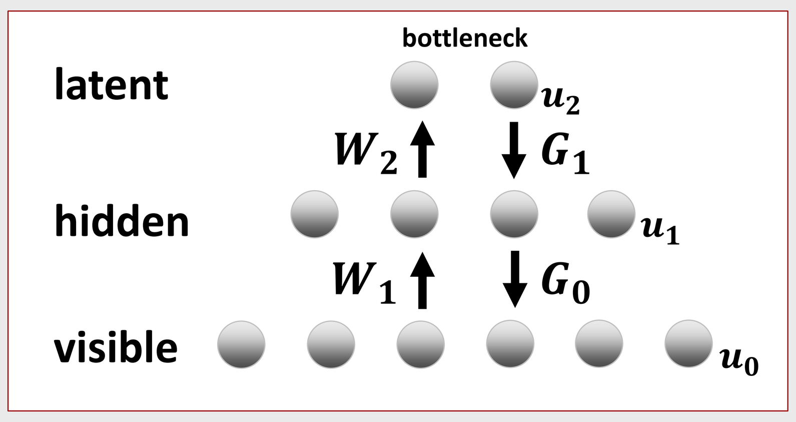


- neuronal dynamics derived from **Euler-Lagrange equations + prospective coding**
- derived neurodynamics: leaky integrators with look-ahead dynamics $\rho_i(t) = r_i(t) + \tau r_i'(t)\dot{u}_i(t) \approx \varphi(u_i(t+\tau))$ **(A)**
- **no phases,** time-continuous backprop **(B)**
- same error interpretation as the dendritic microcircuit model
- see also: **Poster T19** by Kungl, Akos F. et al.





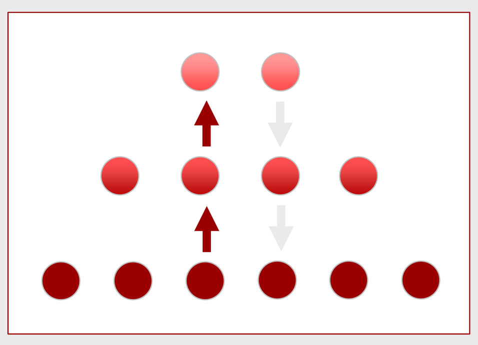
2. Folded autoencoder structure[6]



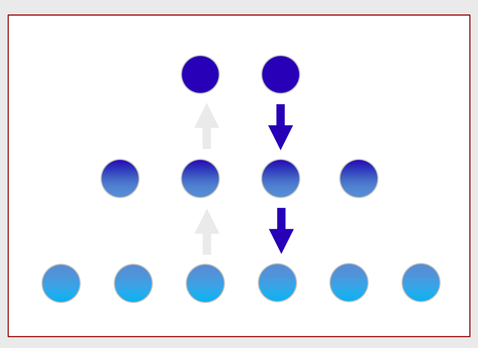
W_i: discriminative / forward weights

G_i: generative / backward weights

Learning useful latent representations through bottleneck, similar to **autoencoders**.



visible input: encoded in latent space through forward weights.



latent input: decode through backward weights to generate data.

→ **happens simultaneously!**

Summary

- Recently, several models demonstrated biologically plausible approximations of backprop
- Here, we extend these to unsupervised learning in a folded autoencoder architecture. **Decoding, encoding and errors are all propagated through the same neurons.**
- **Learning is implemented by dendritic prediction of somatic activity.** Forward and backward weights optimize different quantities, even though formally the plasticity rules are identical.

Also: check out my new publication on Bayesian inference in deterministic spiking networks! :)
Stochasticity from function — Why the Bayesian brain may need no noise, https://doi.org/10.1016/j.neunet.2019.08.002

3. Model and neurophysiological interpretation

Network structure and dynamics condensed in energy function (squared errors):

$$\frac{\gamma}{g_l} E = \frac{\lambda}{2} \sum_i^N \|\boldsymbol{u}_i - \boldsymbol{W}_i \boldsymbol{r}_{i-1}\|^2 + \frac{1-\lambda}{2} \sum_i^N \|\boldsymbol{u}_i - \boldsymbol{G}_i \boldsymbol{r}_{i+1}\|^2 + \beta C$$

λ: gating of forward and backward flow, C: reconstr. error $C = \|\boldsymbol{u}_0 - \boldsymbol{u}_0^{trgt}\|^2$

g_l: leak conductance, $\gamma = 1 + \frac{g_\epsilon}{g_l} \approx 1$

Neurosynaptic dynamics derived as gradient descent:

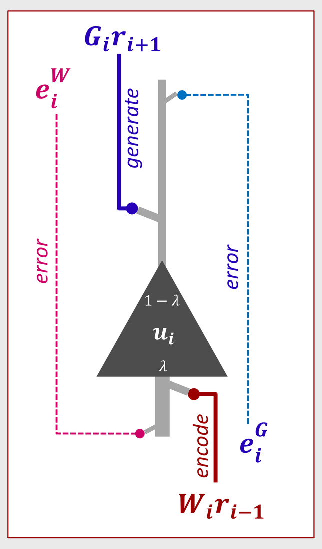
$$c_m \dot{\boldsymbol{u}}_i = -\frac{\partial E}{\partial \boldsymbol{u}_i}, \quad \dot{\boldsymbol{W}}_i = -\eta_W \frac{\partial E}{\partial \boldsymbol{W}_i}, \quad \dot{\boldsymbol{G}}_i = -\eta_G \frac{\partial E}{\partial \boldsymbol{G}_i}.$$

$$\tau \dot{\boldsymbol{u}}_i = \boldsymbol{g}_l^\lambda(\gamma \boldsymbol{W}_i \boldsymbol{r}_{i-1} - \boldsymbol{u}_i) + \boldsymbol{g}_\epsilon^\lambda(\gamma \boldsymbol{e}_i^W - \boldsymbol{u}_i) + \boldsymbol{g}_l^{1-\lambda}(\gamma \boldsymbol{G}_i \boldsymbol{r}_{i+1} - \boldsymbol{u}_i) + \boldsymbol{g}_\epsilon^{1-\lambda}(\gamma \boldsymbol{e}_i^G - \boldsymbol{u}_i)$$

with
$$\boldsymbol{e}_i^W = \frac{\boldsymbol{g}_l}{\boldsymbol{g}_\epsilon} \boldsymbol{r}_i' \cdot \boldsymbol{W}_{i+1}^T(\boldsymbol{u}_{i+1} - \boldsymbol{W}_{i+1}\boldsymbol{r}_i), \quad \boldsymbol{e}_i^G = \frac{\boldsymbol{g}_l}{\boldsymbol{g}_\epsilon} \boldsymbol{r}_i' \cdot \boldsymbol{G}_{i-1}^T(\boldsymbol{u}_{i-1} - \boldsymbol{G}_{i-1}\boldsymbol{r}_i)$$
(prediction errors)

and $\boldsymbol{g}_x^\lambda = \lambda \boldsymbol{g}_x$, $\boldsymbol{g}_x^{1-\lambda} = (1-\lambda)\boldsymbol{g}_x$

→ **5-compartment model** with soma and four dendritic branches.



Plasticity:

$$\dot{\boldsymbol{W}}_i \propto [\boldsymbol{u}_i - \boldsymbol{W}_i \boldsymbol{r}_{i-1}]\boldsymbol{r}_{i-1}^T$$
$$\dot{\boldsymbol{G}}_i \propto [\boldsymbol{u}_i - \boldsymbol{G}_i \boldsymbol{r}_{i+1}]\boldsymbol{r}_{i+1}^T$$

Both plasticity rules can be interpreted as Urbanczik-Senn type[7] rules:
learning is driven by the dendritic prediction of somatic activity.

For edge cases ($\lambda = 0$ or $\lambda = 1$), the plasticity approximates error backpropagation, e.g., for $\lambda = 0$ we get $\dot{\boldsymbol{G}}_i \propto \boldsymbol{e}_i^G \cdot \boldsymbol{r}_{i+1}^T$ and $\boldsymbol{e}_i^G = \frac{\boldsymbol{g}_l}{\boldsymbol{g}_\epsilon} \boldsymbol{r}_i' \cdot \boldsymbol{G}_{i-1}^T \boldsymbol{e}_{i-1}^G$.

4. One learning rule, two optimizations

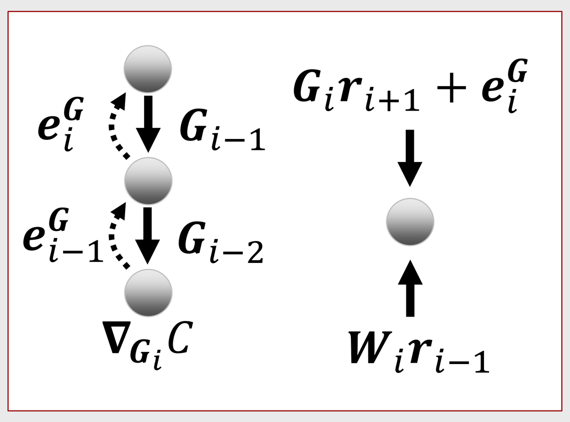
Using the solution of stationary neurodynamics as well as choosing $\lambda \ll 1, \lambda > 0$ **the plasticity rules can be rewritten as:**

$$\dot{\boldsymbol{G}}_i \propto -(1-\lambda)^2 \cdot \nabla_{\boldsymbol{G}_i} C$$

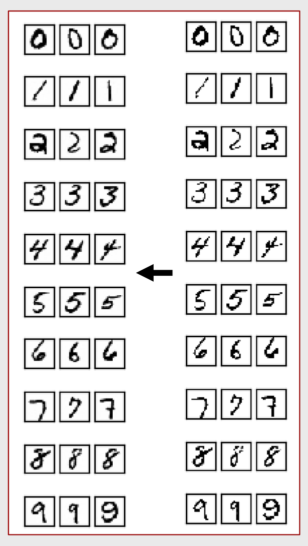
reduce cost function via backprop, e.g., with dendritic model

$$\dot{\boldsymbol{W}}_i \propto -\lambda(1-\lambda) \cdot \nabla_{\boldsymbol{W}_i} \left\| \boldsymbol{G}_i \boldsymbol{r}_{i+1} + \boldsymbol{e}_i^G - \boldsymbol{W}_i \boldsymbol{r}_{i-1} \right\|^2$$

learn to match generative input at each layer

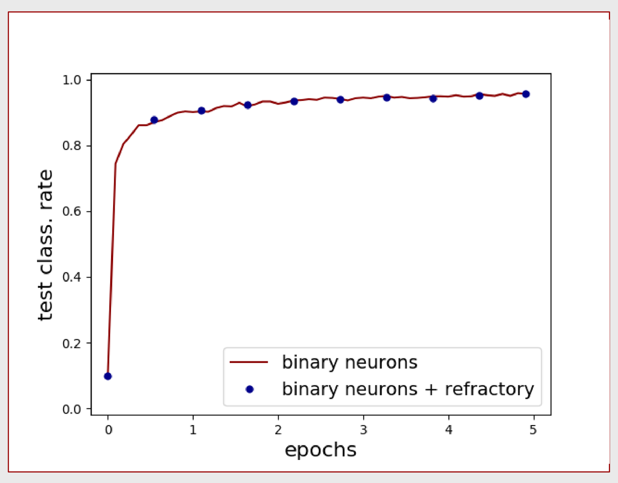


Test by encoding with discriminative and decoding with generative path:



5. Outlook

- **Bidirectional learning** by adding cost function in latent layer → currently work in progress!
- **Spiking neuron models?** Initial results for classification of MNIST images with stochastic binary neurons and refractory period of 3ms.



[1]Whittington, James C., & Bogacz, Rafal. **An approximation of the error backpropagation algorithm in a predictive coding network with local Hebbian synaptic plasticity**. Neural computation (2017).
[2]Guerguiev, Jordan, et al. **Towards deep learning with segregated dendrites**. Elife (2017).
[3]Sacramento, João, et al. **Dendritic cortical microcircuits approximate the backpropagation algorithm**. NeurIPS (2018).
[4]Scellier, Benjamin, & Bengio, Yoshua. **Equilibrium propagation: Bridging the gap between energy-based models and backpropagation**. Frontiers in computational neuroscience (2017).
[5]Dold, Dominik, et al. **Lagrangian dynamics of dendritic microcircuits enables real-time backpropagation of errors**. Cosyne Abstracts (2019).
[6]Seung, H. S. **Learning continuous attractors in recurrent networks** in Advances in neural information processing systems (1998).
[7]Urbanczik, Robert, & Senn, Walter. **Learning by the dendritic prediction of somatic spiking**. Neuron 81.3 (2014).

This work has received funding from the **Manfred Stärk Foundation** and the **European Union Horizon 2020 Framework Programme** under grant agreement 720270 and 785907 (HBP). Calculations using GPUs were performed on **UBELIX** (http://www.id.unibe.ch/hpc), the HPC cluster of the University of Bern. Furthermore, this work was supported by the state of Baden-Württemberg through **bwHPC** and the **German Research Foundation (DFG)** through grant no INST 39/963-1 FUGG.

Email contact: dodo@kip.uni-heidelberg.de