Unsupervised discovery of neural sequences in large-scale recordings

Emily Mackevicius

Andrew Bahle

Alex Williams

Shijie Gu

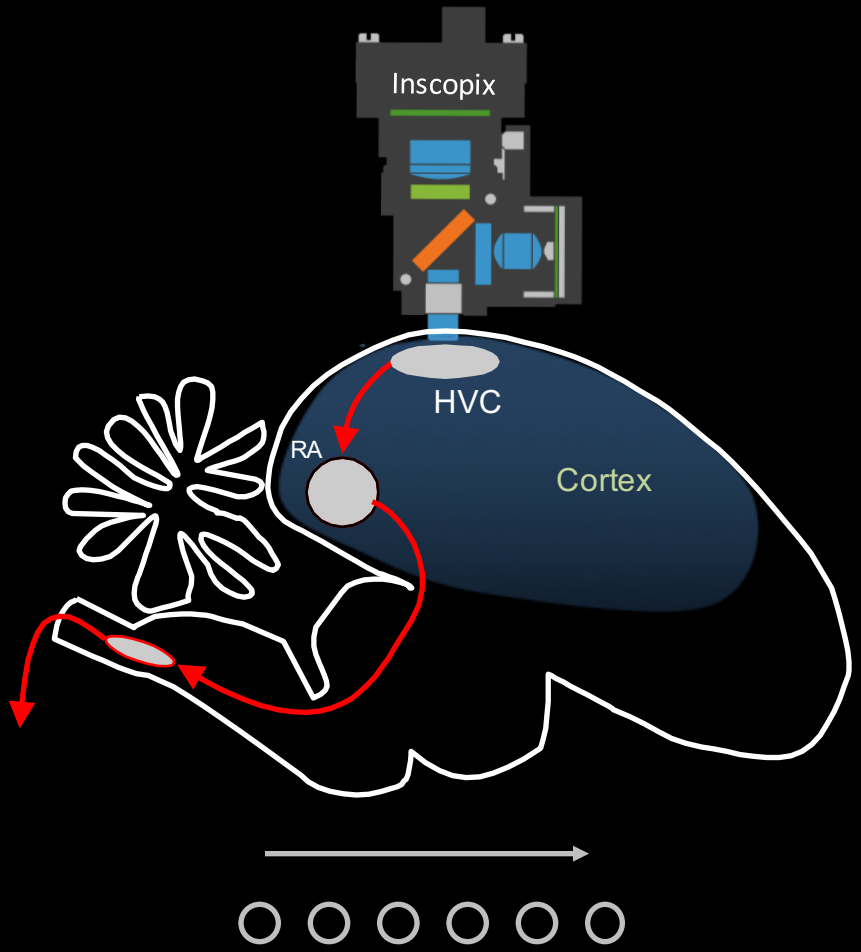
Natasha Denissenko

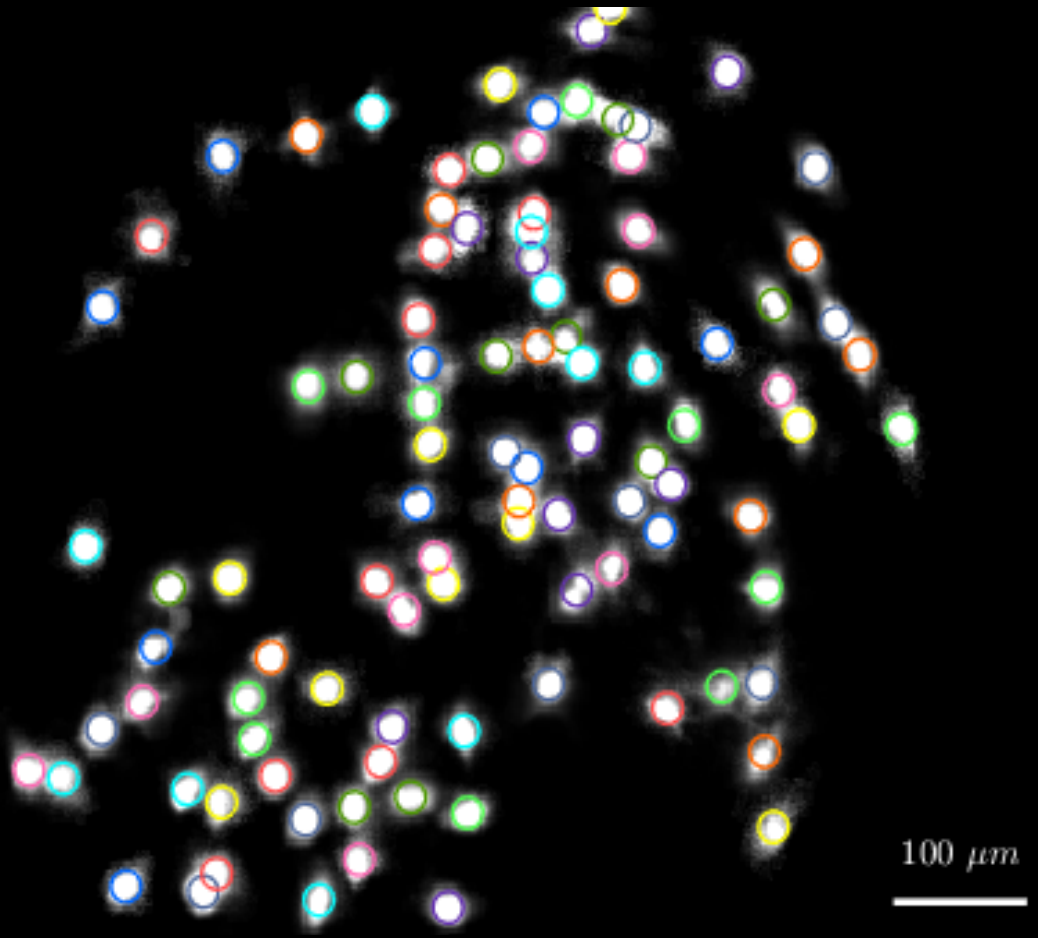
Mark Goldman

Michale Fee

github.com/Feelab/seqNMF

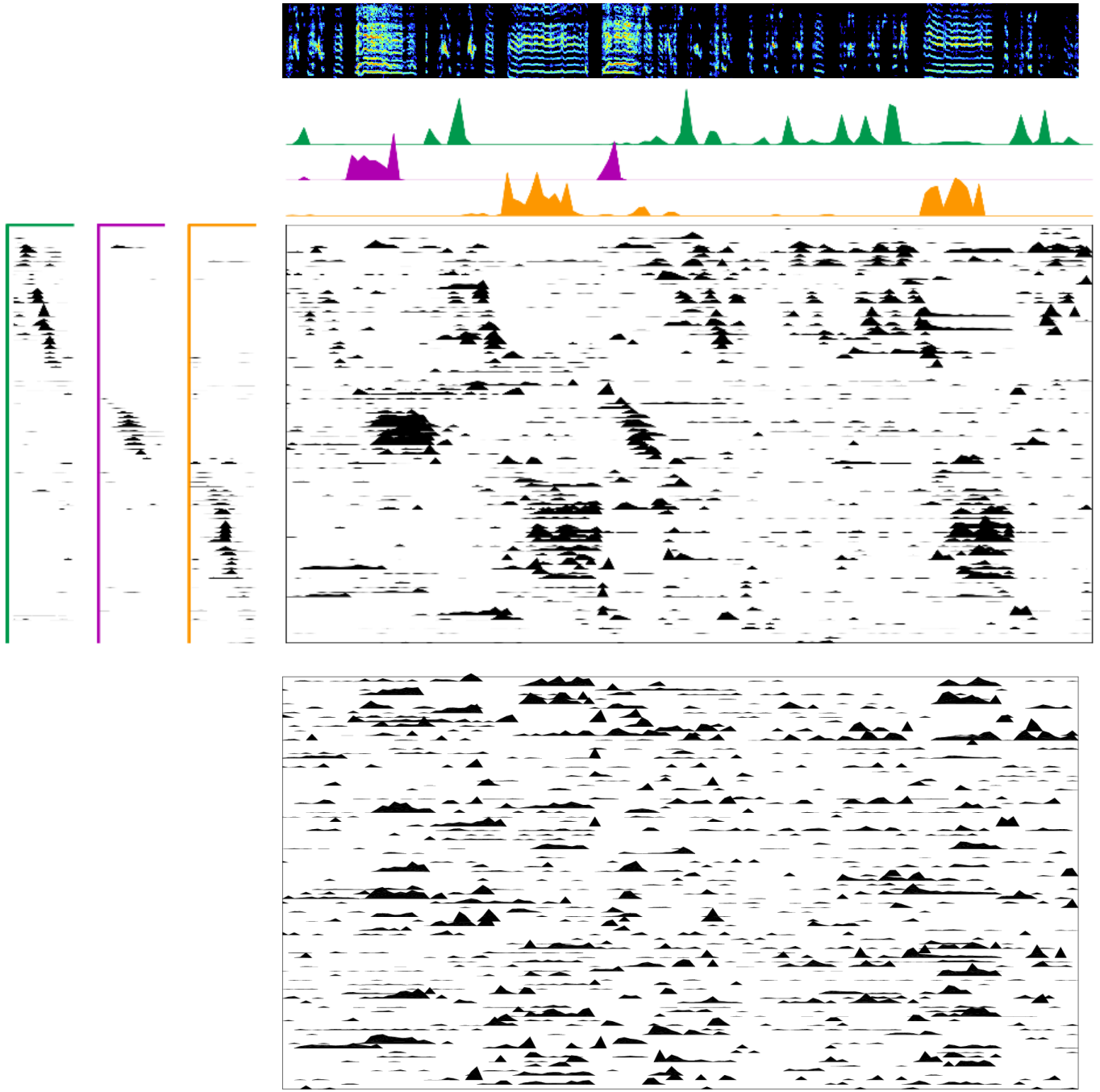
https://www.biorxiv.org/content/early/2018/03/02/273128


Inscopix
HVC
RA
Cortex

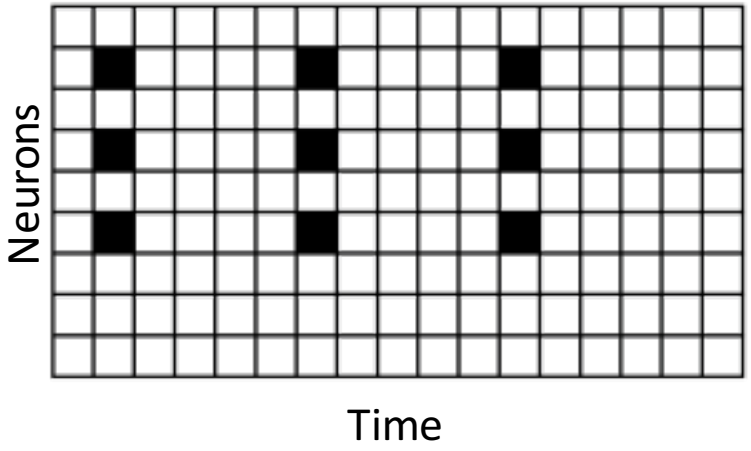


CNMF_E cell extraction (PC Zhou, Paninski lab, with Shijie Gu)

Neurons
(re-sorted)



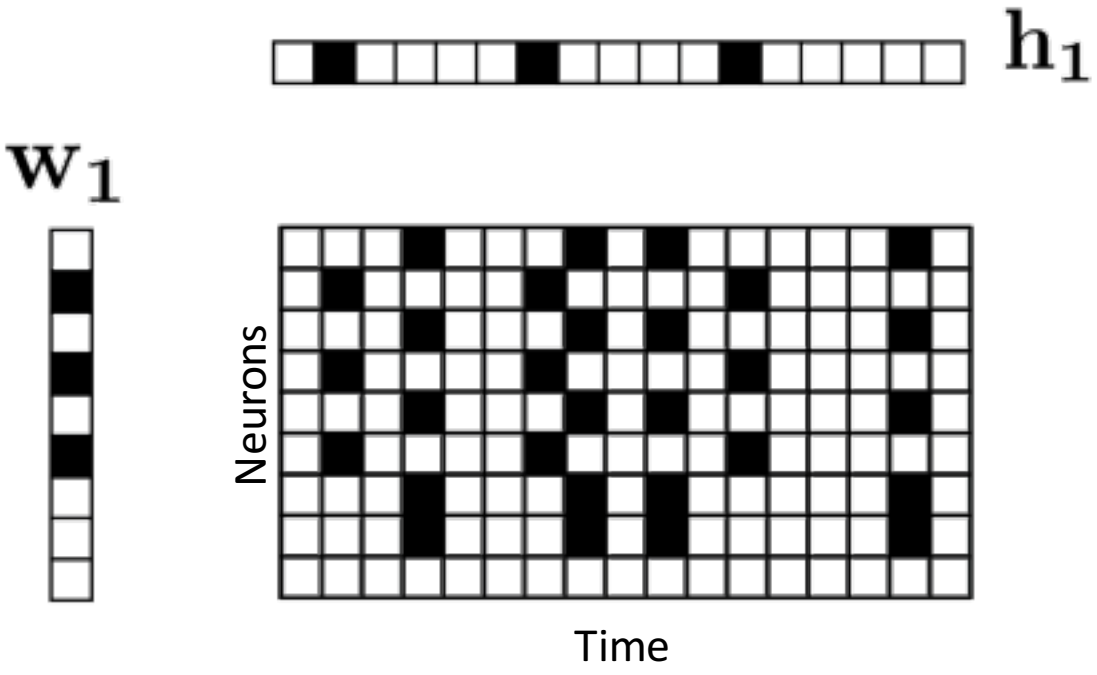
Non-negative matrix factorization (NMF)



Non-negative matrix factorization (NMF)

(like PCA/SVD, but factors must be positive)

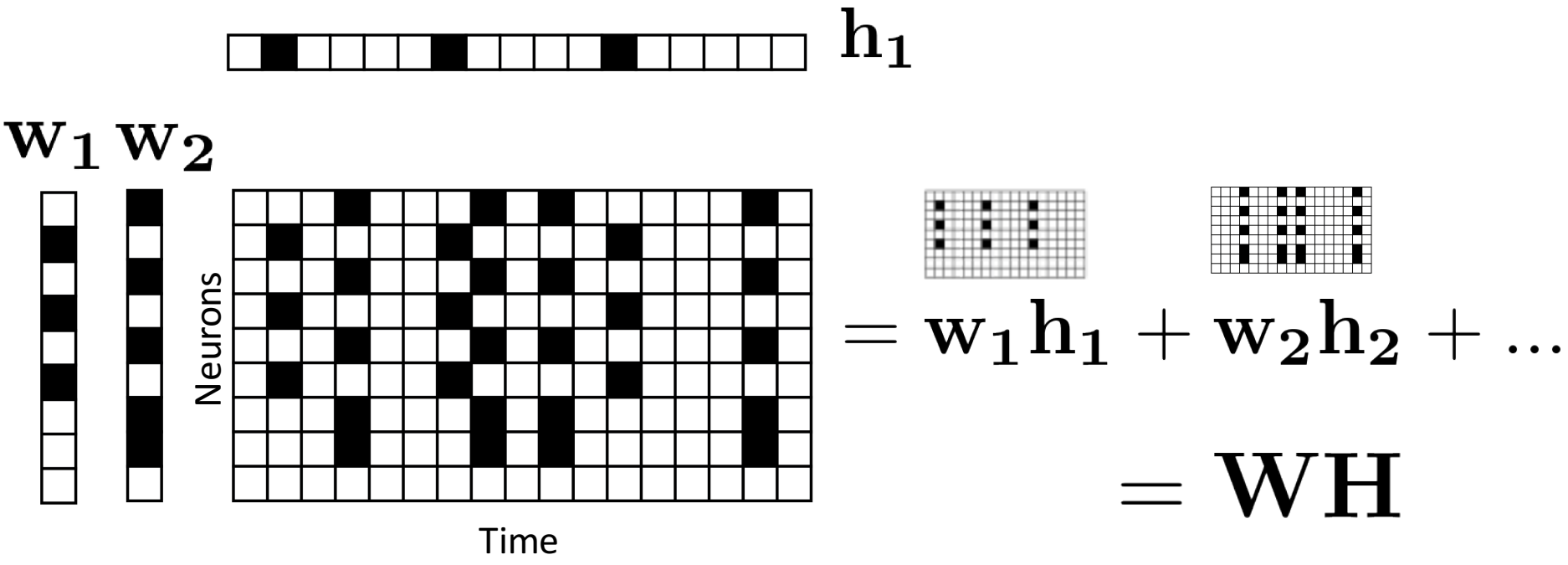
Non-negative matrix factorization (NMF)



Non-negative matrix factorization (NMF)

(like PCA/SVD, but factors must be positive)

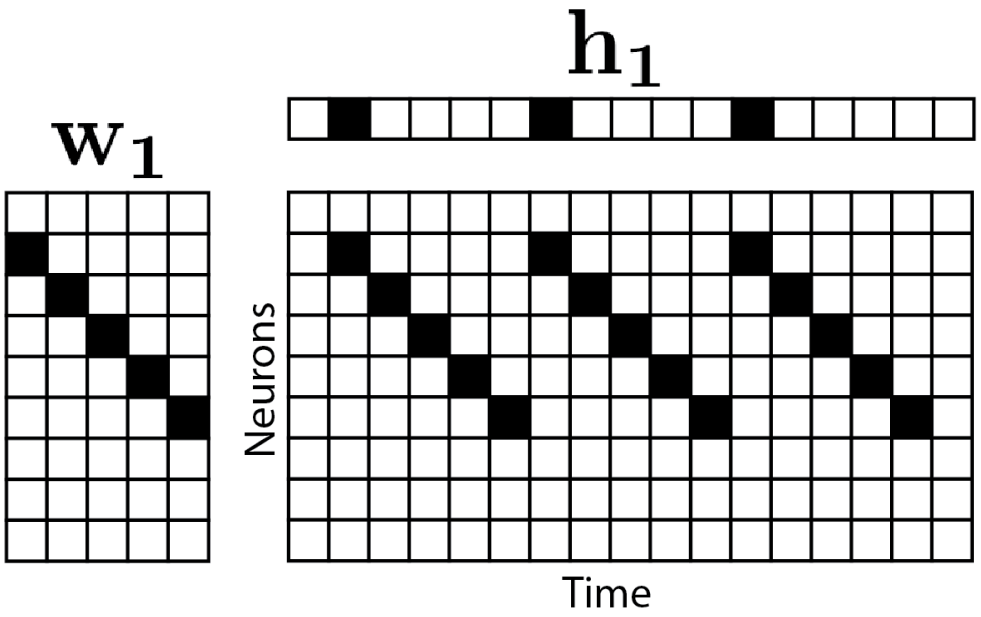
Non-negative matrix factorization (NMF)



Non-negative matrix factorization (NMF)

(like PCA/SVD, but factors must be positive)

Convolutional NMF



$$+ \mathbf{w_2} \circledast \mathbf{h_2} + \cdots$$

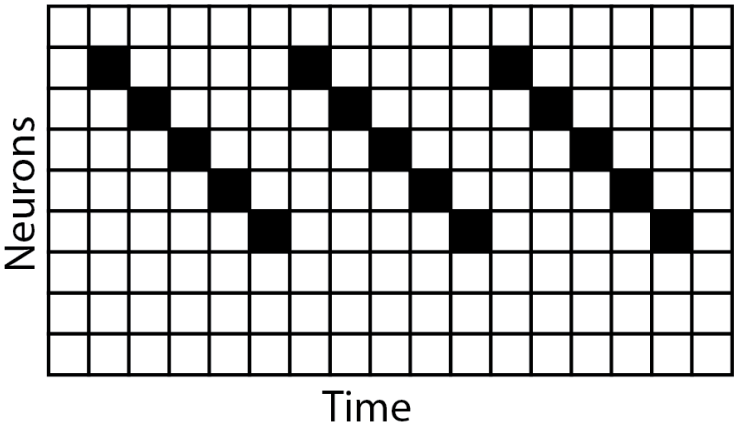
$$= \mathbf{W} \circledast \mathbf{H}$$

Easy to fit

$$(\widetilde{\mathbf{W}}, \widetilde{\mathbf{H}}) = \arg\min_{\mathbf{W},\mathbf{H}} \left(||\mathbf{X} - \mathbf{W} \circledast \mathbf{H}||_F^2 \right)$$

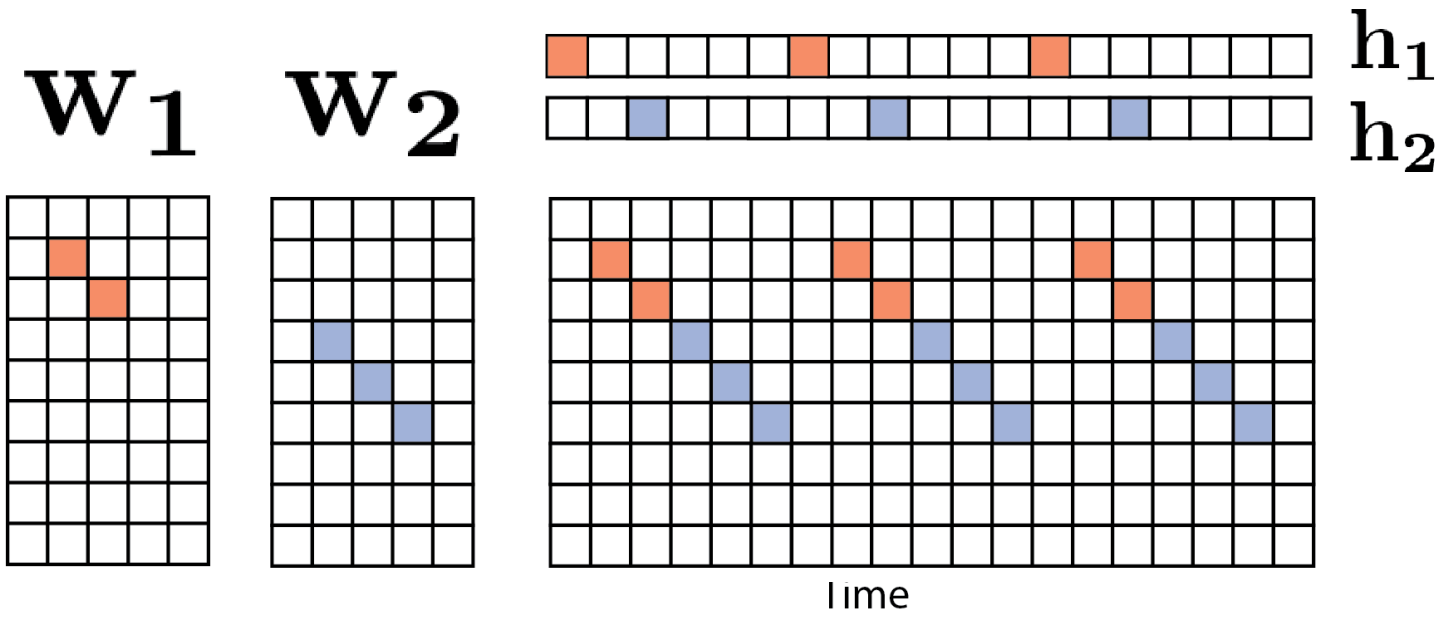
Smaragdis, 2004, 2007

In practice … redundant factors



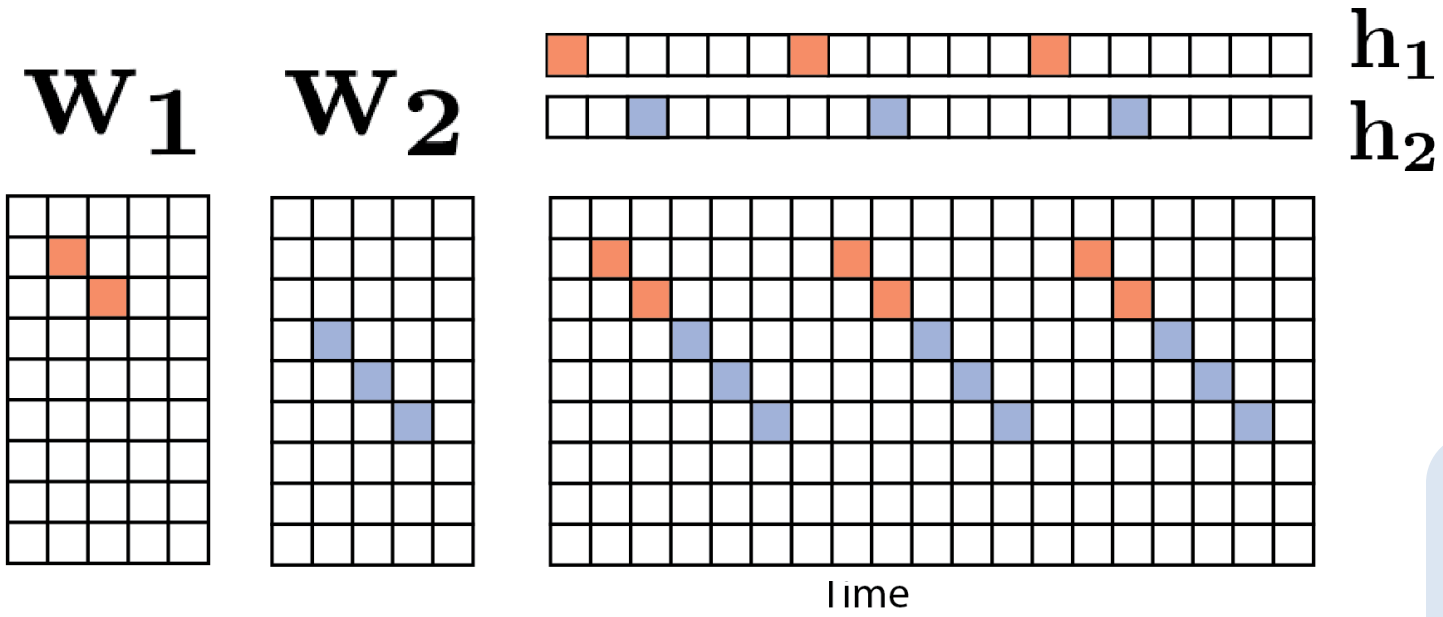
$$(\widetilde{\mathbf{W}}, \widetilde{\mathbf{H}}) = \arg\min_{\mathbf{W},\mathbf{H}} \left(||\mathbf{X} - \mathbf{W} \circledast \mathbf{H}||_F^2 \right)$$

In practice … redundant factors



$$(\widetilde{\mathbf{W}}, \widetilde{\mathbf{H}}) = \arg\min_{\mathbf{W},\mathbf{H}} \left(||\mathbf{X} - \mathbf{W} \circledast \mathbf{H}||_F^2 \right)$$

In practice … redundant factors

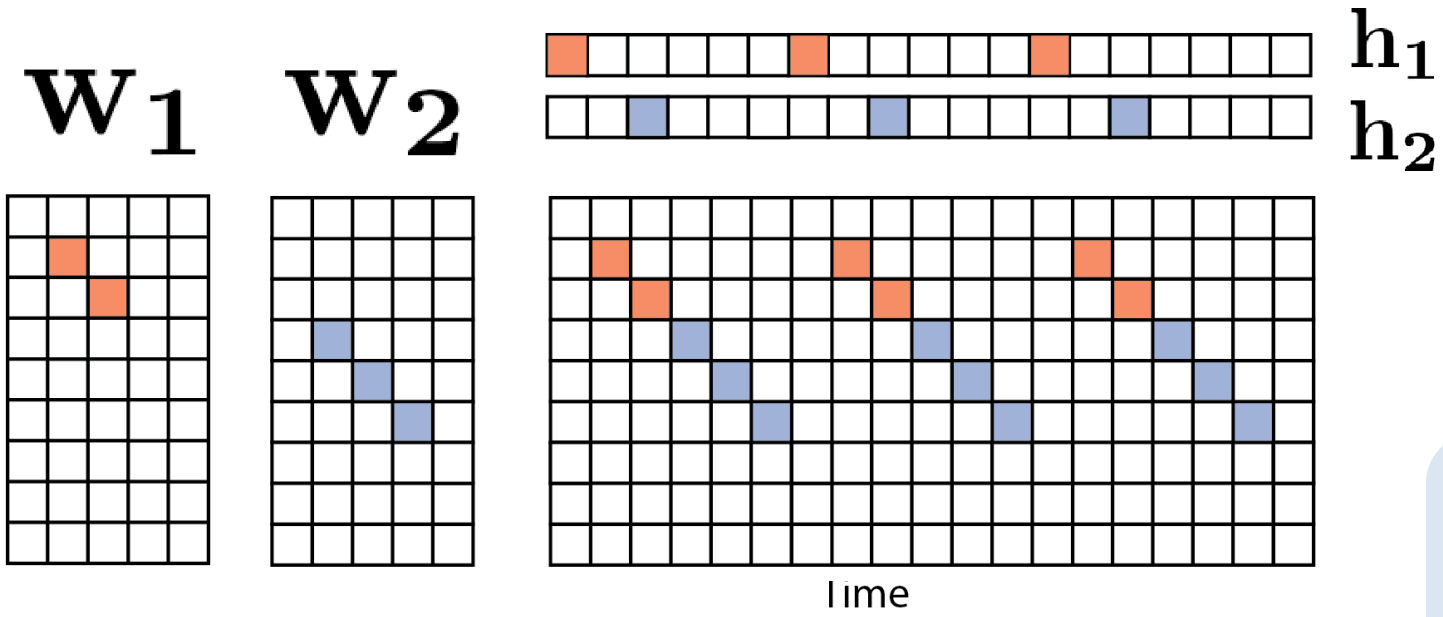


Cost term punishes correlations

$$(\widetilde{\mathbf{W}}, \widetilde{\mathbf{H}}) = \arg\min_{\mathbf{W},\mathbf{H}} \left(||\mathbf{X} - \mathbf{W} \circledast \mathbf{H}||_F^2 + \overbrace{\lambda \mathcal{R}} \right)$$

$$\mathcal{R} = ||\mathbf{H}\mathbf{H}^\top||$$

In practice … redundant factors

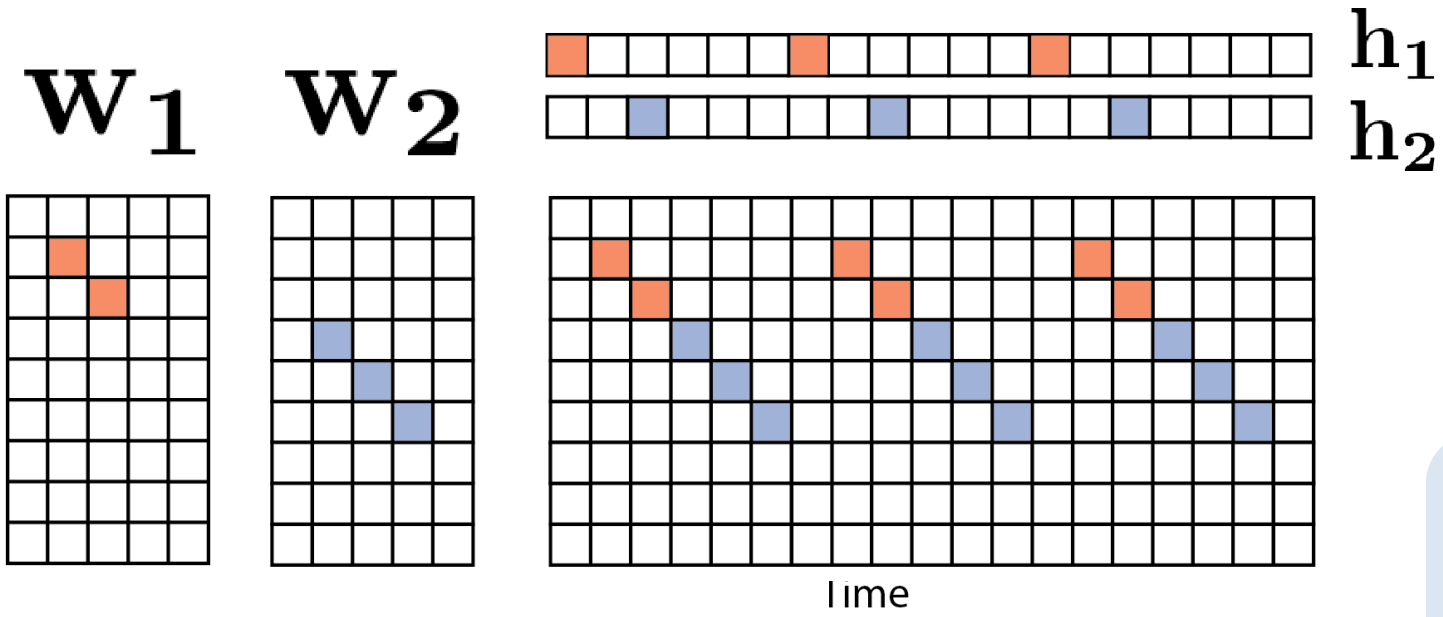


Cost term punishes correlations

$$(\widetilde{\mathbf{W}}, \widetilde{\mathbf{H}}) = \arg\min_{\mathbf{W},\mathbf{H}} \left(||\mathbf{X} - \mathbf{W} \circledast \mathbf{H}||_F^2 + \overbrace{\lambda \mathcal{R}} \right)$$

$$\mathcal{R} = ||\mathbf{H}\mathbf{S}\mathbf{H}^\top||$$

In practice … redundant factors

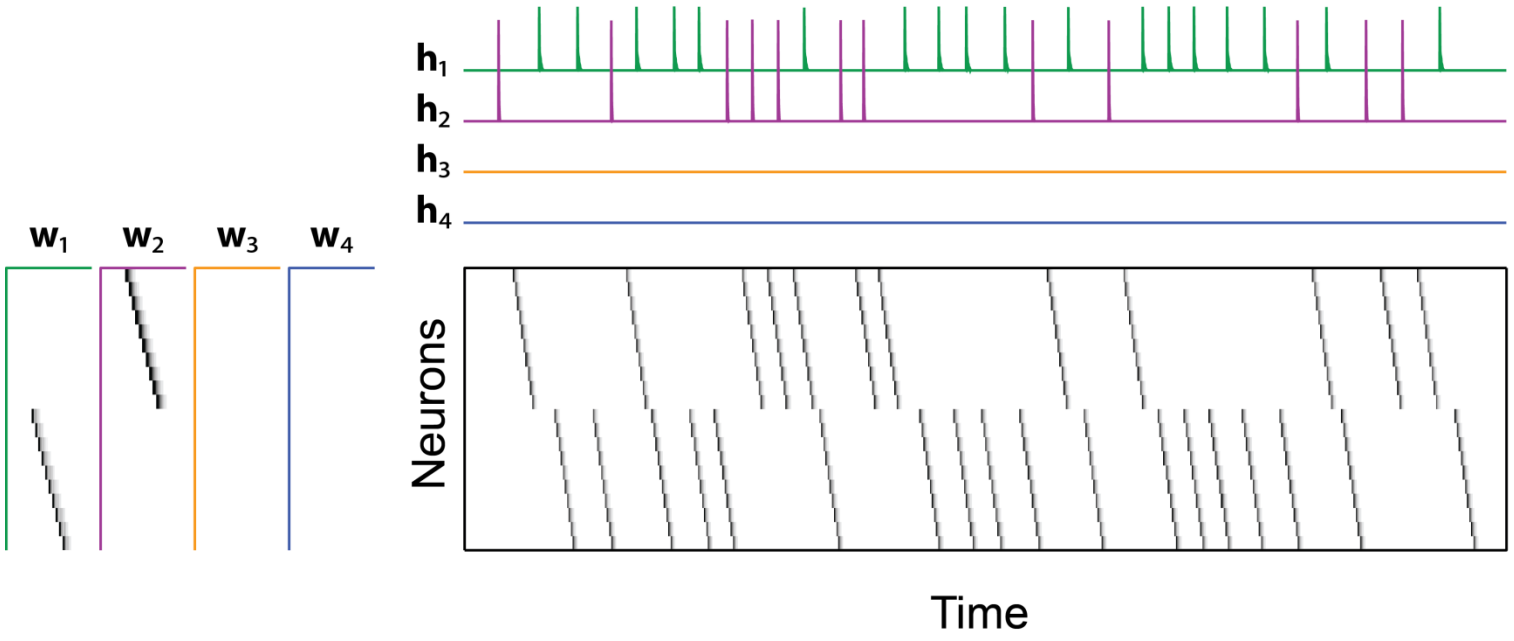


Cost term punishes correlations

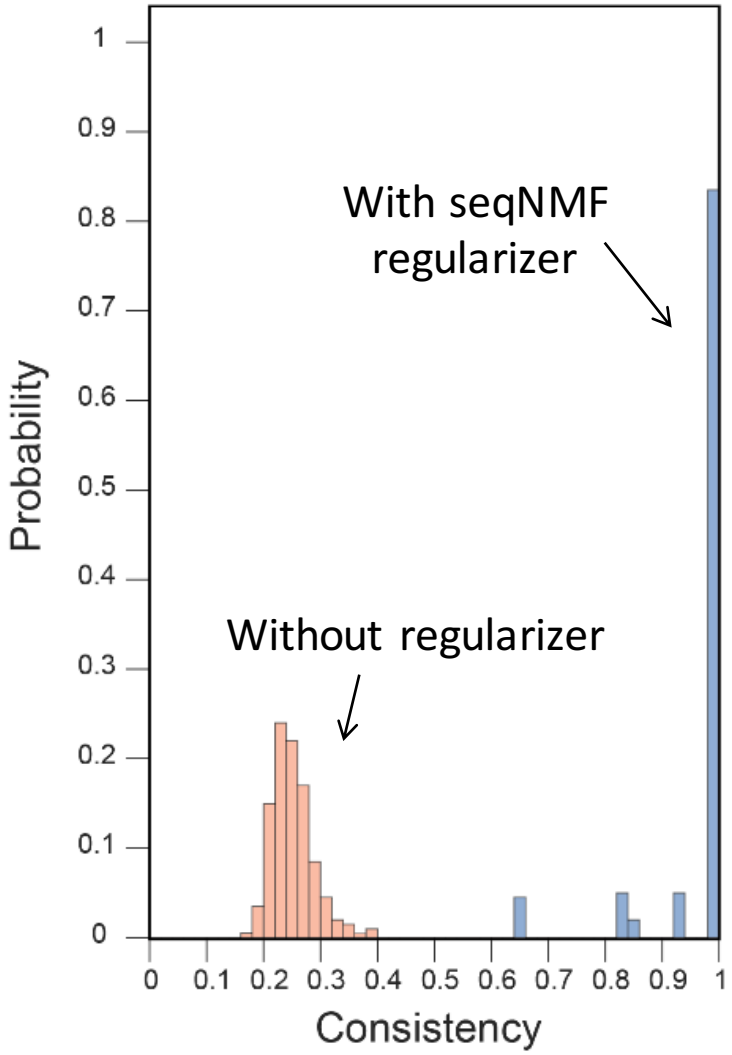
$$(\widetilde{\mathbf{W}}, \widetilde{\mathbf{H}}) = \arg\min_{\mathbf{W},\mathbf{H}} \left(||\mathbf{X} - \mathbf{W} \circledast \mathbf{H}||_F^2 + \overbrace{\lambda \mathcal{R}} \right)$$

$$\mathcal{R} = ||\mathbf{W} \overset{\top}{\circledast} \mathbf{X}\mathbf{S}\mathbf{H}^\top||$$

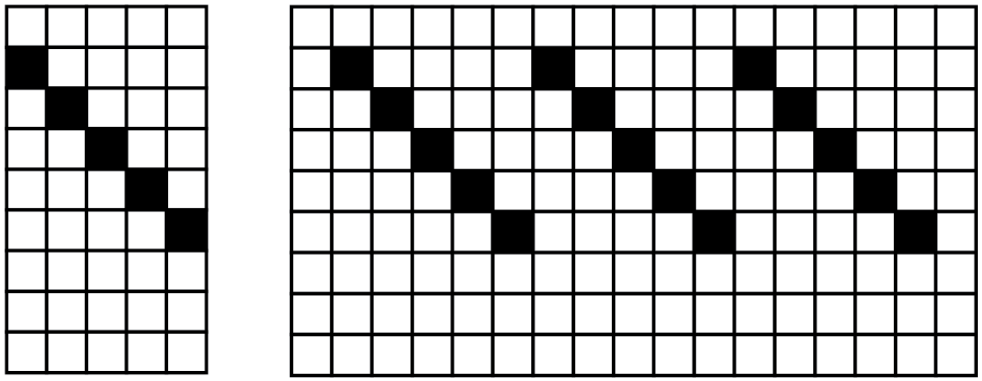
SeqNMF robustly discovers factors in simulated data



SeqNMF factorizations are highly consistent compared to CNMF

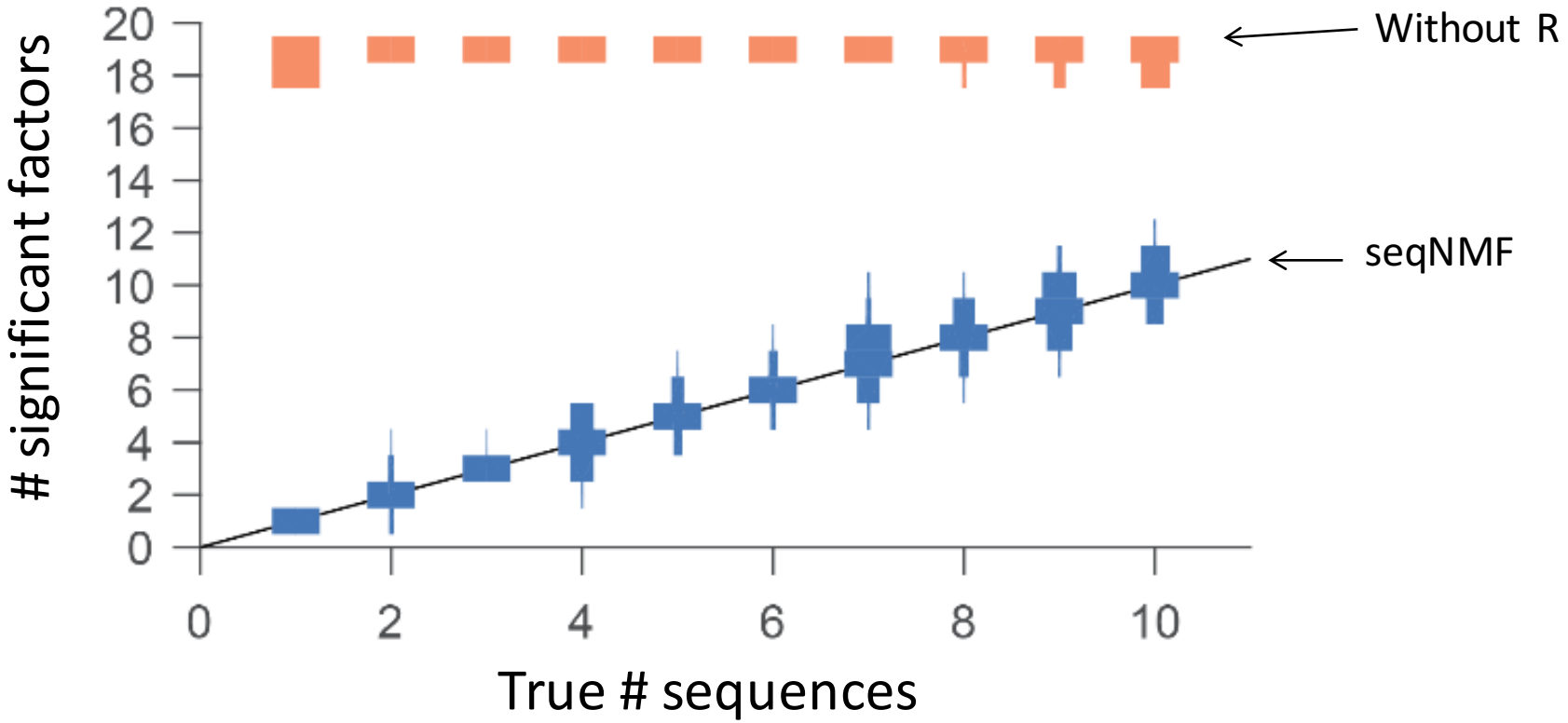


Testing significance of each factor on held-out data

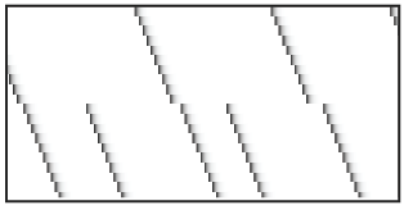


Many moments of high overlap, compared to null

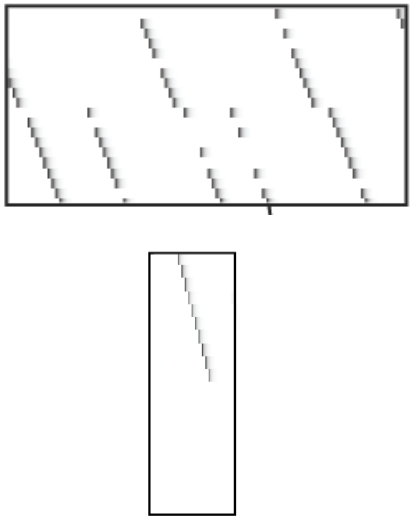
SeqNMF discovers the correct number of sequences



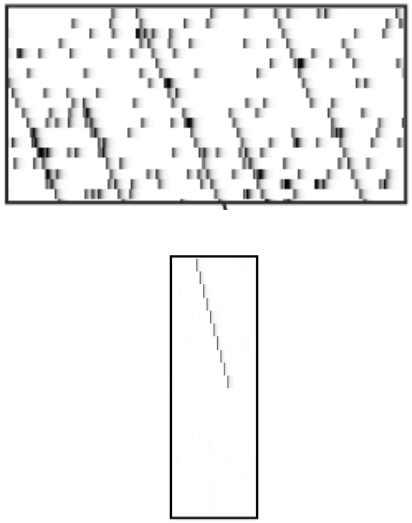
SeqNMF is robust to noise



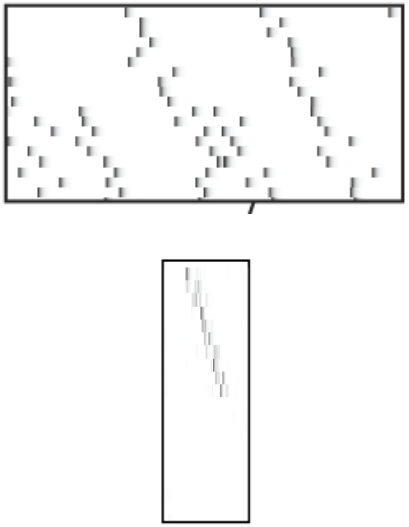
Probabilistic participation



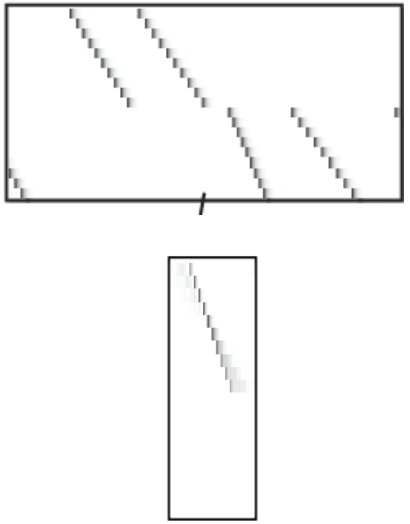
Additive noise



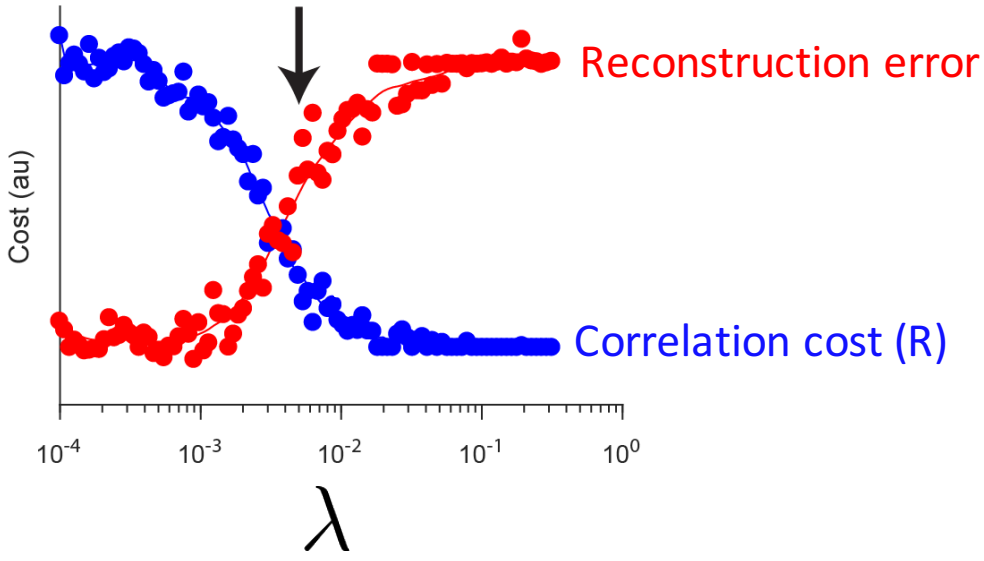
Temporal jitter



Temporal warping



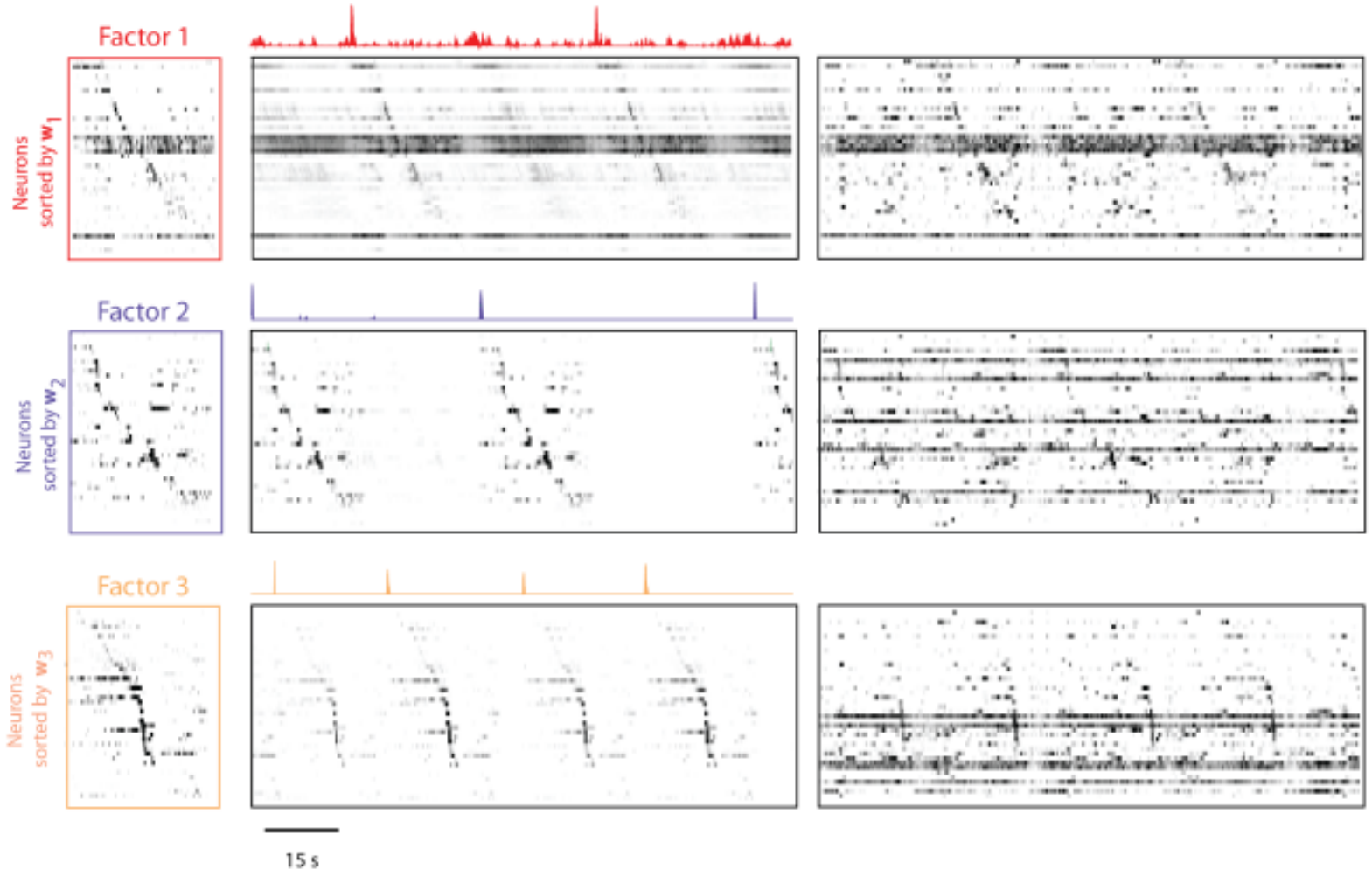
Method to choose lambda



Cost term punishes correlations

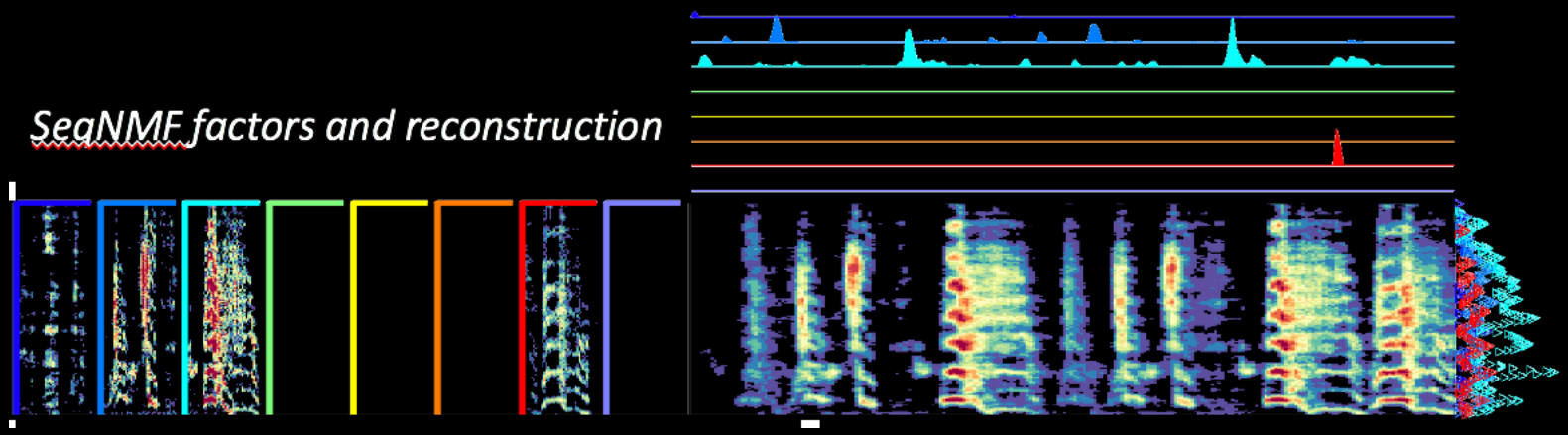
$$(\widetilde{\mathbf{W}}, \widetilde{\mathbf{H}}) = \arg\min_{\mathbf{W},\mathbf{H}} \left(||\mathbf{X} - \mathbf{W} \circledast \mathbf{H}||_F^2 + \lambda \mathcal{R} \right)$$

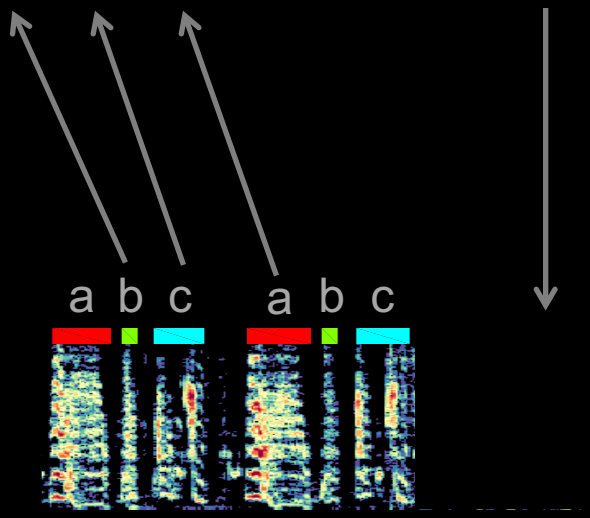
SeqNMF extracts hippocampal sequences without reference to the behavior



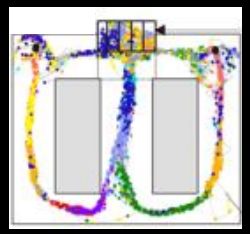
Pastalkova et al. 2008, Buszaki lab

SeqNMF is broadly applicable to many high-dimensional datasets





Hand-labeled song syllables, Okubo 2015



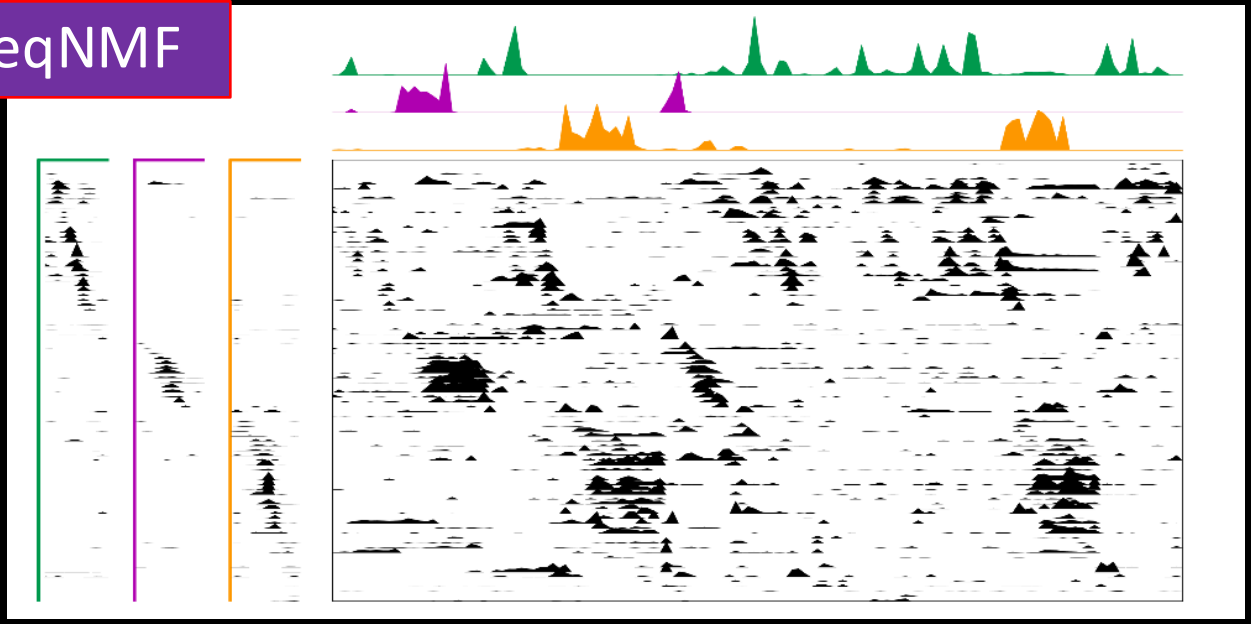
Hippocampal sequences
from
Pastalkova et al,
Buszaki lab

. . .

SeqNMF: GO FIND SEQUENCES!

github.com/Feelab/seqNMF

elm@mit.edu
abahle@mit.edu



Emily Mackevicius

Andrew Bahle

Alex Williams

Shijie Gu

Natasha Denissenko

Mark Goldman

Michale Fee

https://www.biorxiv.org/content/early/2018/03/02/273128