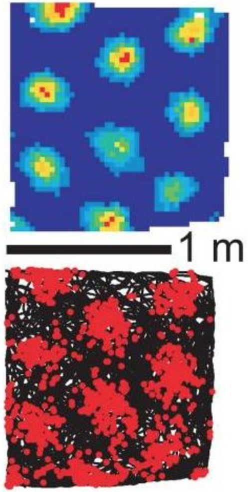
A unified theory for the origin of grid cells through the lens of pattern formation

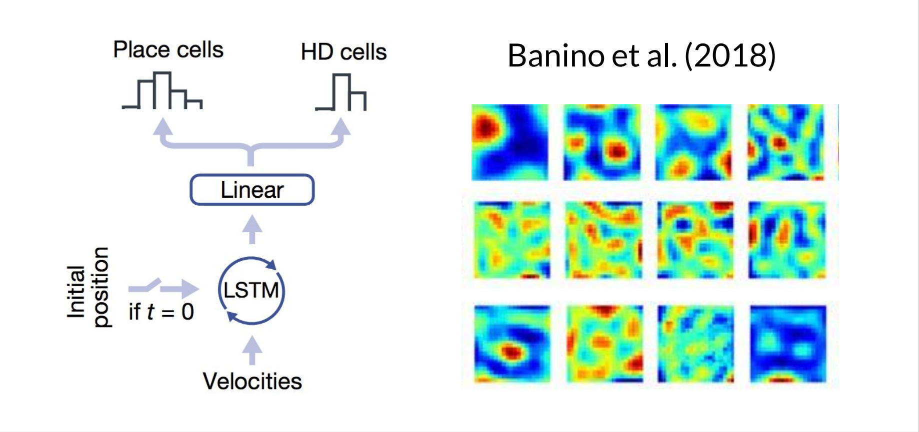
Ben Sorscher*, Gabriel C. Mel*, Surya Ganguli, Sam Ocko

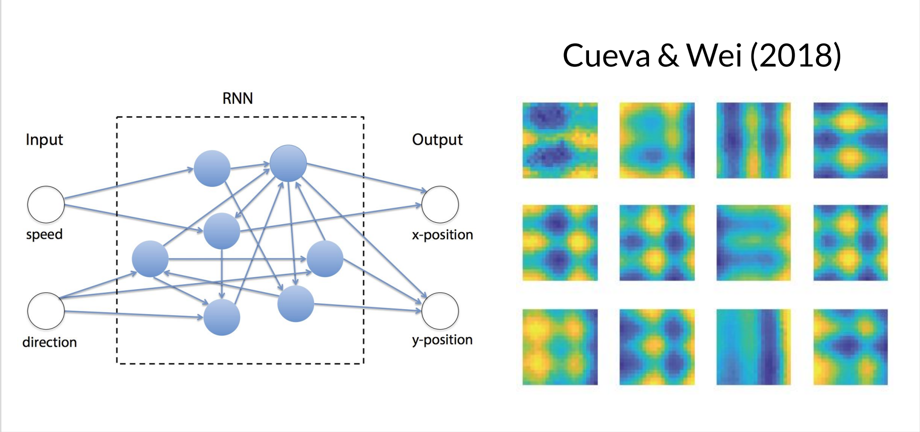
Grid cells

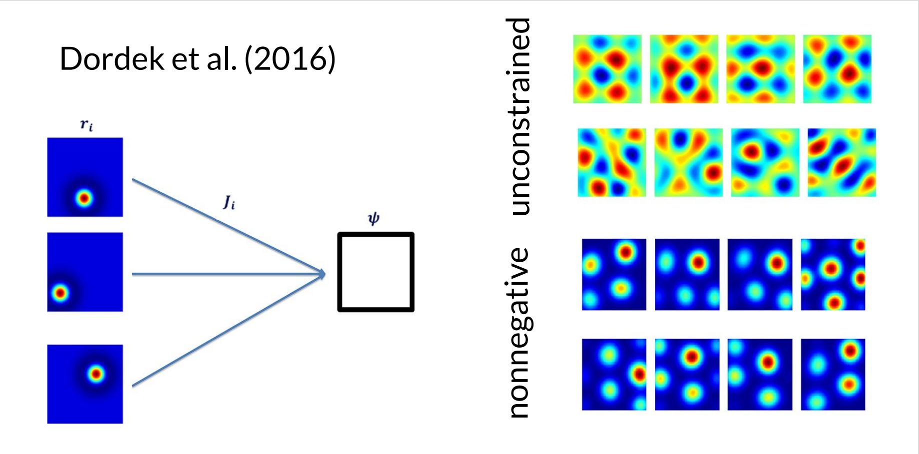


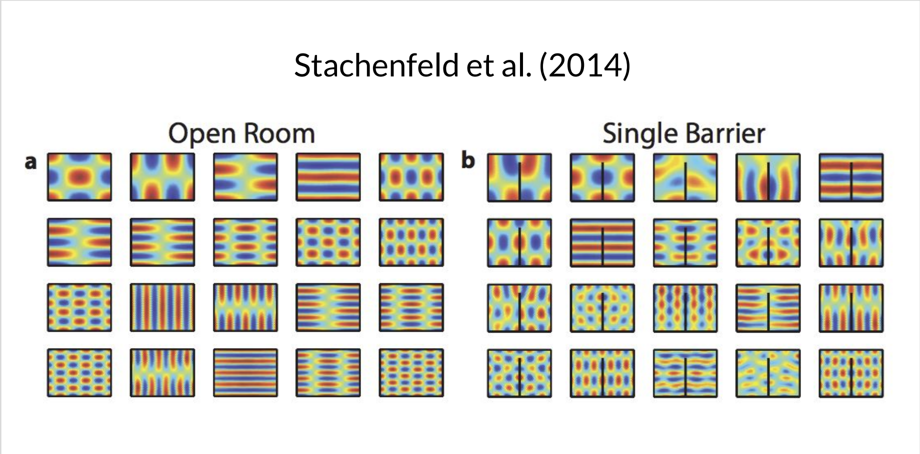
Krupic et al. (2012), Derdikman (2014)

Trained neural networks learn grid patterns

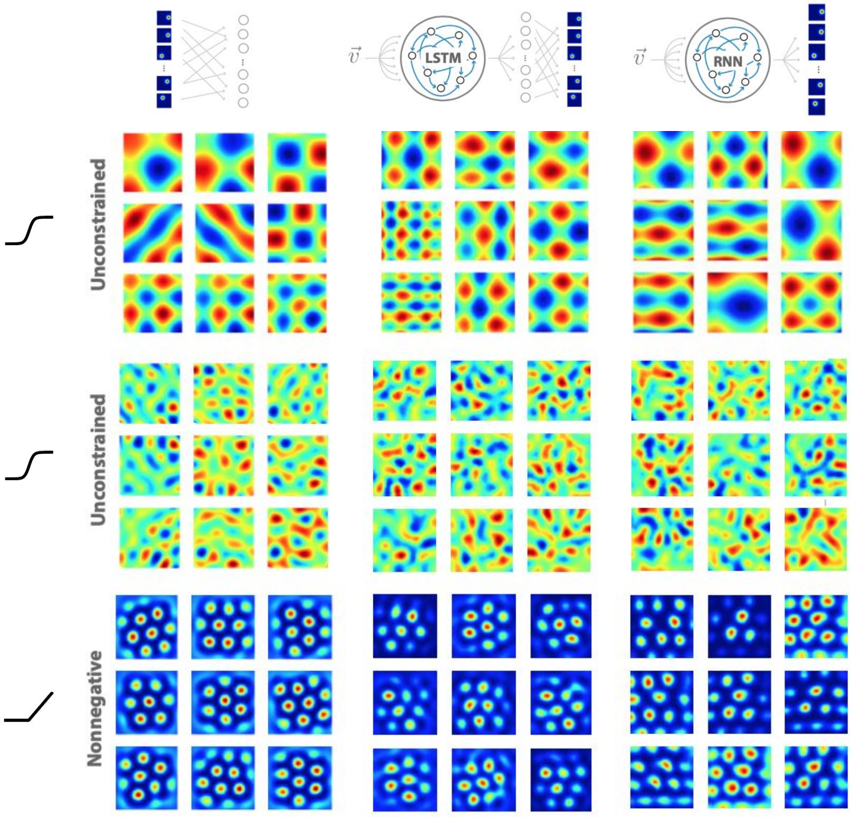








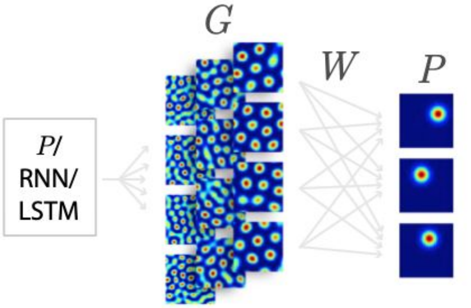
Trained neural networks learn grid patterns



1. Why are the optimal maps grids?

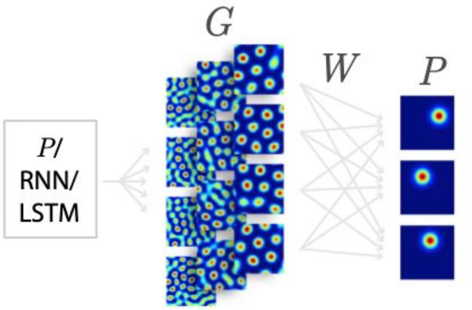
2. What determines the optimal grid type - square, amorphous, or hexagonal?

Gradient descent as a *pattern forming dynamics*

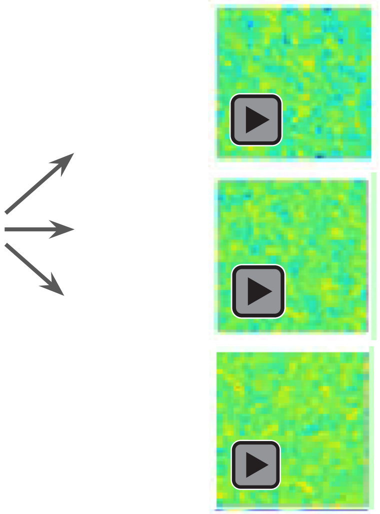


$$\min_{W,G} \|P - WG\|^2$$

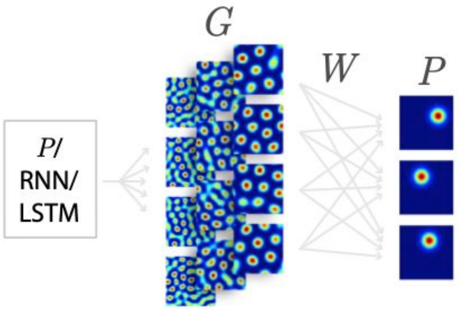
Gradient descent as a *pattern forming dynamics*



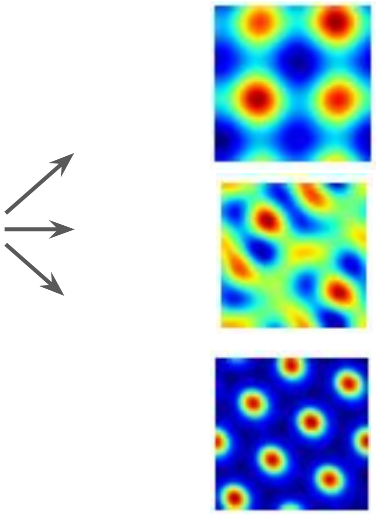
$$\min_{W,G} \|P - WG\|^2$$



Gradient descent as a *pattern forming dynamics*

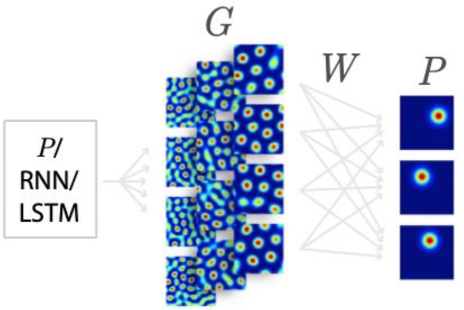


$$\min_{W,G} \|P - WG\|^2$$

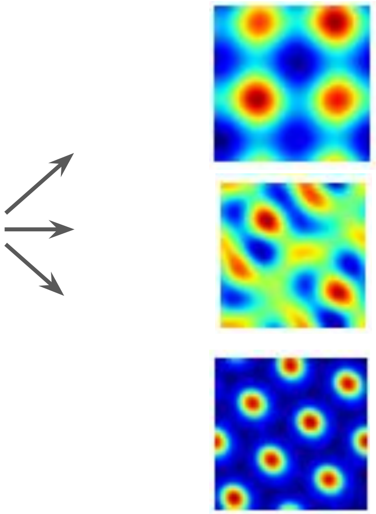


1. Why are the optimal maps grids?

Gradient descent as a *pattern forming dynamics*

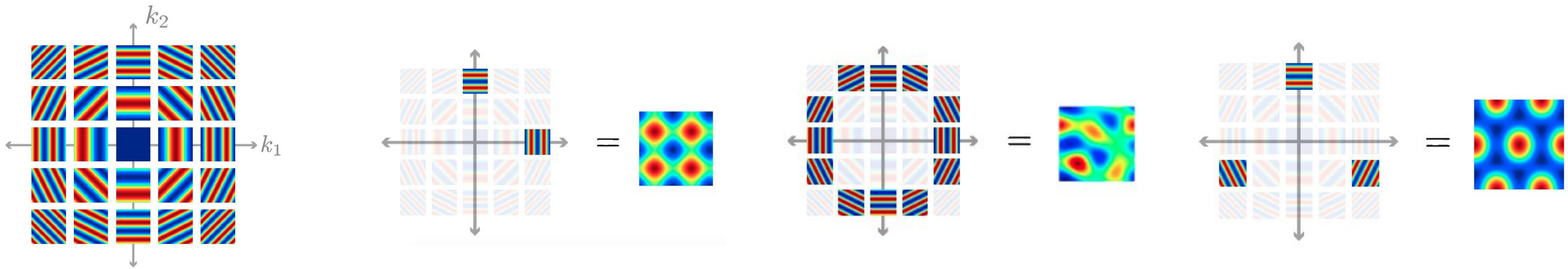


$$\min_{W,G} \|P - WG\|^2$$



1. Why are the optimal maps grids?

Translation invariance => Fourier modes



1. Why are the optimal maps grids?

    Translation invariance => Fourier modes

2. What determines the optimal grid type - square, amorphous, or hexagonal?

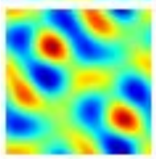
1. Why are the optimal maps grids?

    Translation invariance => Fourier modes

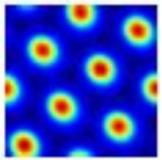
2. What determines the optimal grid type - square, amorphous, or hexagonal?

Nonnegativity yields hexagonal grids

$$\min_{W,G} \|P - WG\|^2$$

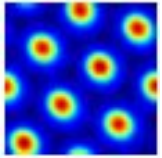


$$\min_{W,G} \|P - WG\|^2 + \sigma(G)$$



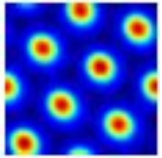
Nonnegativity yields hexagonal grids

$$\min_{W,G} \|P - WG\|^2 + \sigma(G)$$



Nonnegativity yields hexagonal grids

$$\min_{W,G} \|P - WG\|^2 + \sigma(G)$$

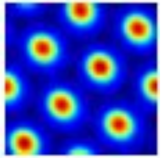


Taylor expand constraint σ

$$\sigma(G) \approx \sigma_0 + \sigma_1 G + \frac{1}{2}\sigma_2 G^2 + \frac{1}{6}\sigma_3 G^3 + \cdots$$

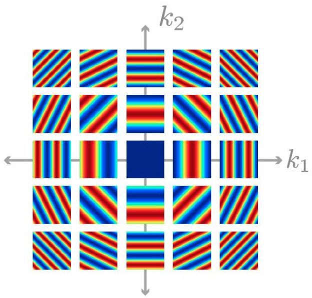
Nonnegativity yields hexagonal grids

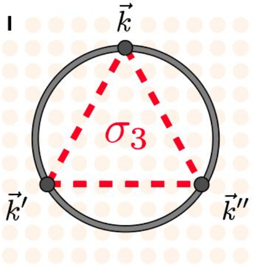
$$\min_{W,G} \|P - WG\|^2 + \sigma(G)$$



Taylor expand constraint σ

$$\sigma(G) \approx \sigma_0 + \sigma_1 G + \frac{1}{2}\sigma_2 G^2 + \frac{1}{6}\sigma_3 G^3 + \cdots$$

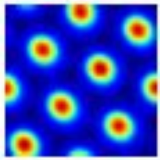




“Three-body interaction” between stripes 60° apart

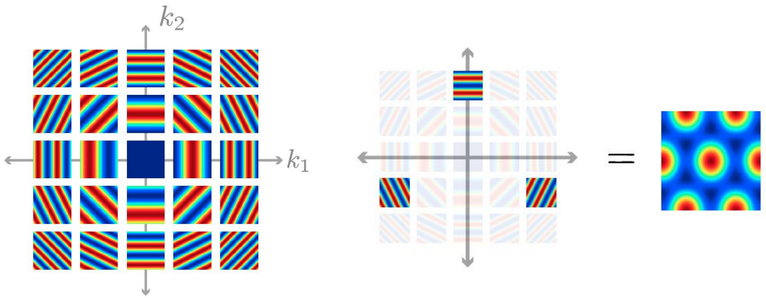
Nonnegativity yields hexagonal grids

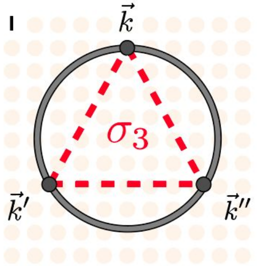
$$\min_{W,G} \|P - WG\|^2 + \sigma(G)$$



Taylor expand constraint σ

$$\sigma(G) \approx \sigma_0 + \sigma_1 G + \frac{1}{2}\sigma_2 G^2 + \frac{1}{6}\sigma_3 G^3 + \cdots$$

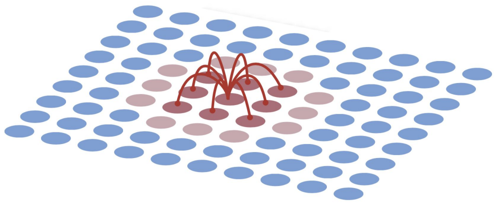




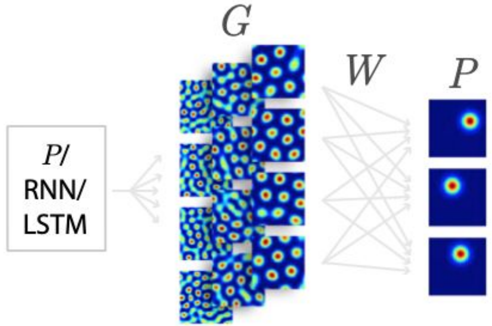
“Three-body interaction” between stripes 60º apart

Unifying mechanistic and normative models

Grid cell model RNNs



Normative encoding models

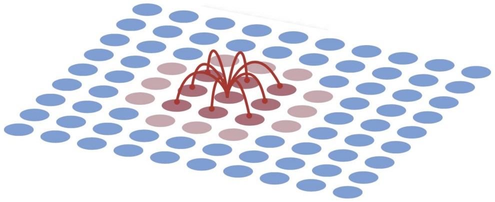


Skaggs et al. (1995)
Zhang (1996)
Fuhs and Touretzky (2006)
Burak and Fiete (2009)

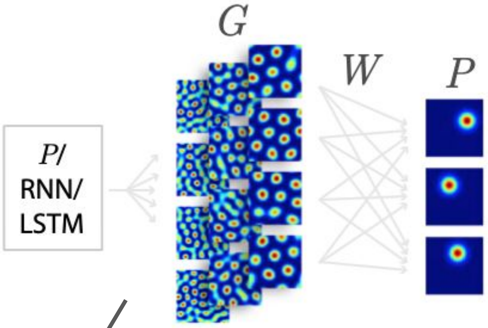
Banino et al. (2018)
Cueva & Wei (2018)
Dordek et al. (2016)
Stachenfeld et al. (2014)

Unifying mechanistic and normative models

Grid cell model RNNs

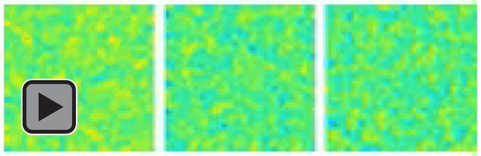


Normative encoding models



Activity dynamics

Gradient descent



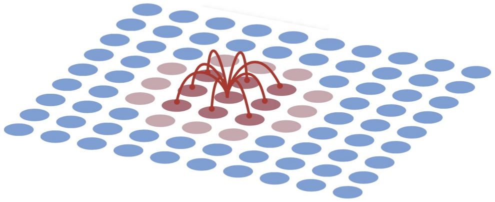
Pattern forming dynamics

Skaggs et al. (1995)
Zhang (1996)
Fuhs and Touretzky (2006)
Burak and Fiete (2009)

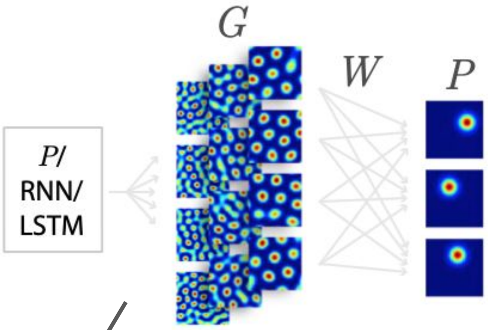
Banino et al. (2018)
Cueva & Wei (2018)
Dordek et al. (2016)
Stachenfeld et al. (2014)

Unifying mechanistic and normative models

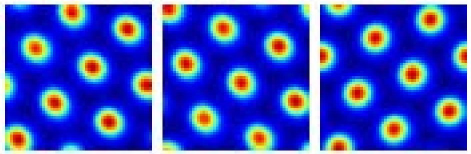
Grid cell model RNNs



Normative encoding models



Activity dynamics = Gradient descent



Pattern forming dynamics

Skaggs et al. (1995)
Zhang (1996)
Fuhs and Touretzky (2006)
Burak and Fiete (2009)

Banino et al. (2018)
Cueva & Wei (2018)
Dordek et al. (2016)
Stachenfeld et al. (2014)