SAMPLING-BASED REPRESENTATION OF UNCERTAINTY
time is of essence

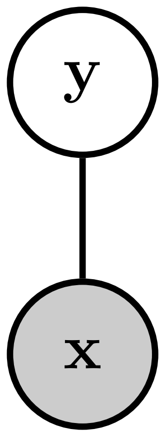
MÁTÉ LENGYEL

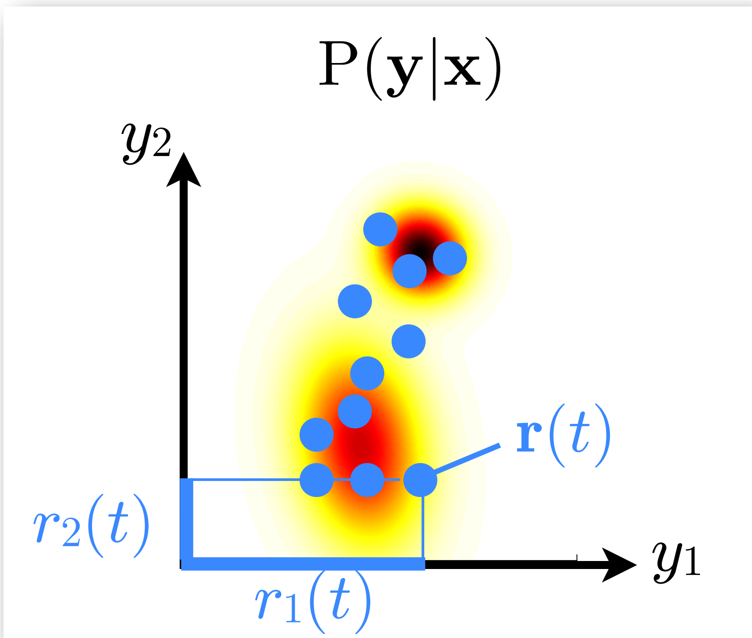
Computational and Biological Learning Lab
Department of Engineering
University of Cambridge

A SIMPLE TAXONOMY OF PROBABILISTIC REPRESENTATIONS

probability distribution

spatio-temporal neural activity patterns

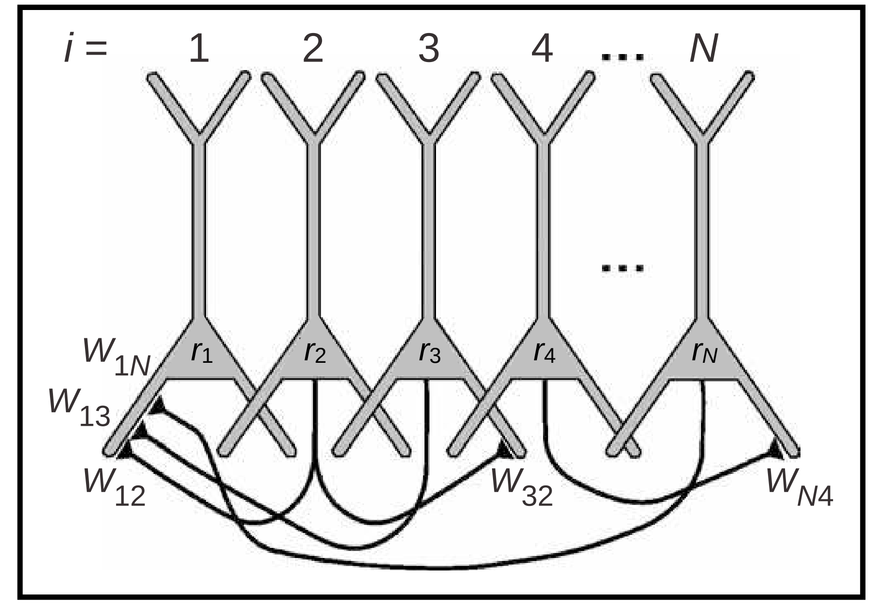


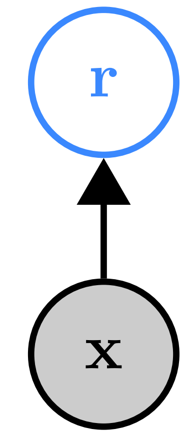


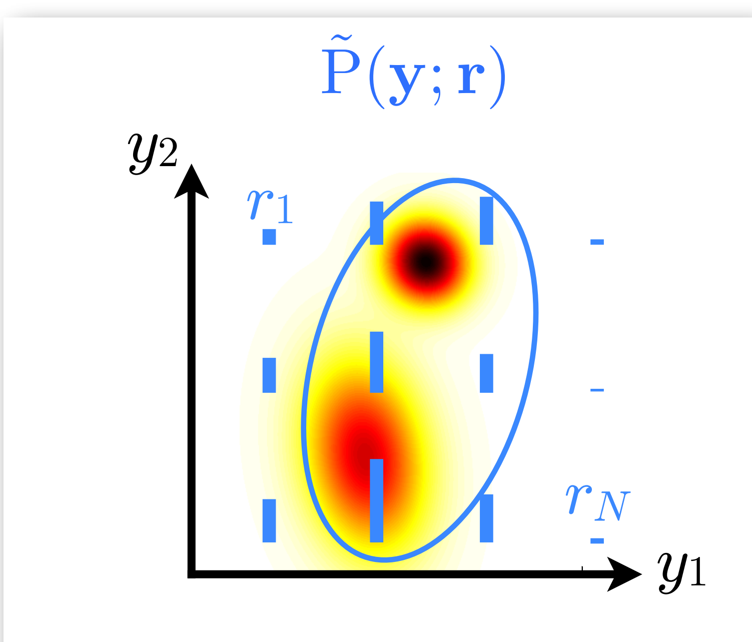
sampling-based

$$\mathbf{r} \sim \mathrm{P}(\mathbf{y} = \mathbf{r}|\mathbf{x})$$

$\mathbf{r}(t)$





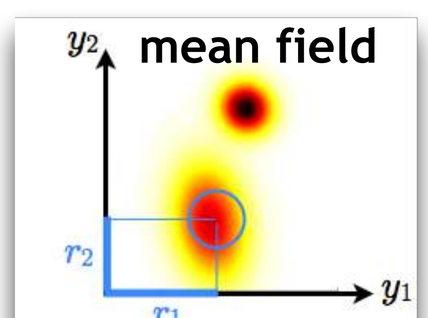


parametric

$$\mathbf{r} = \operatorname{argmin} \mathrm{KL}\left[\tilde{\mathrm{P}}(\mathbf{y};\mathbf{r}) \parallel \mathrm{P}(\mathbf{y}|\mathbf{x})\right]$$ instantaneous

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{r} = -\nabla_{\mathbf{r}} \mathrm{KL}\left[\tilde{\mathrm{P}}(\mathbf{y};\mathbf{r}) \parallel \mathrm{P}(\mathbf{y}|\mathbf{x})\right]$$ iterative

→ **probabilistic population codes (product of experts)**

→ 

→ **predictive coding**

NEURAL REPRESENTATIONS OF UNCERTAINTY

	sampling	mean-field	prob. pop. code	
neurons represent	variables	variables	parameters	
neurons / variable	1	1	many (~100—1000)	too many?
distributions are represented	by iterative sampling	by iterative dynamics / instantaneously	instantaneously	too slow!?
correlations (limiting factor)	✓ (time)	✗	✓ (neurons)	
cue combination	✓	✓	✓	
marginalisation	✓	✓	✓?	
dynamics	stochastic	deterministic	deterministic	
neural variability for computation	useful	harmful	harmful	robustness?
learning	✓	✓?	?	
stimulus-dependent noise correlations	✓	✗	✗	

A PROTO-MODEL: GAUSSIAN SCALE MIXTURE

Wainwright & Simoncelli 2000, Schwartz & Simoncelli 2001, Coen-Cagli et al 2012

visual cortical activities $\mathbf{u} \sim \mathrm{P}(\mathbf{u}|\mathbf{x})$

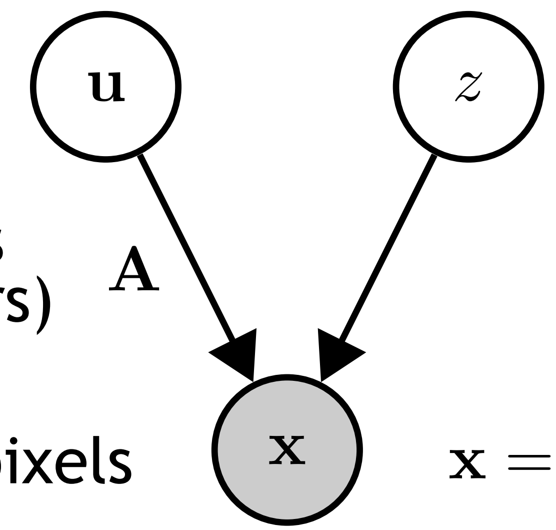
visual features intensities

contrast level

$\mathbf{u} \sim \mathcal{N}(\mathbf{0}, \mathbf{C})$

$z \sim \mathrm{Gamma}(k, \theta)$

visual features
(eg. Gabor filters)



pixels

$\mathbf{x} = z\,\mathbf{A}\,\mathbf{u} + \sigma_{\mathrm{x}}\,\boldsymbol{\epsilon} \qquad \boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$

PLAN OF THE TALK

❖ **neural circuit models**

analog variables
speed
E-I networks

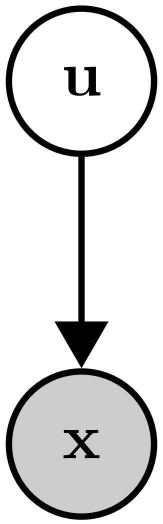
❖ **empirical evidence** (almost model-free)

neural: evoked-spontaneous activity → József Fiser's talk yesterday
behavioural: role of time

NEURAL NETWORK DYNAMICS: A SIMPLE CASE STUDY

Hennequin et al, NIPS 2014

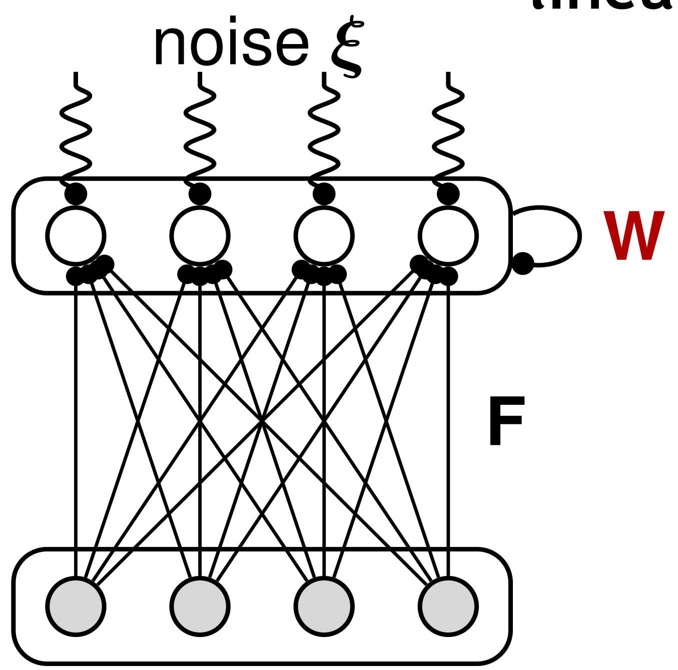
generative model: factor analysis



$$\mathbf{u} \sim \mathcal{N}(\mathbf{0}, \mathbf{C})$$

$$\mathbf{x} = \mathbf{A}\,\mathbf{u} + \sigma_{\mathrm{x}}\,\boldsymbol{\epsilon}$$

$$\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$$

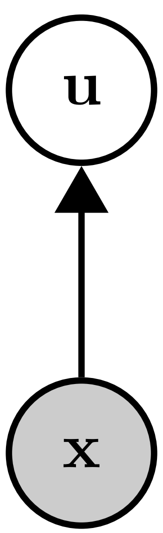


linear network dynamics

$$\mathrm{d}\mathbf{u} = \frac{\mathrm{d}t}{\tau_{\mathrm{m}}}\left(-\mathbf{u} + \mathbf{W}\,\mathbf{u} + \mathbf{F}\,\mathbf{x}\right) + \; + \sigma_{\xi}\sqrt{\frac{2}{\tau_{\mathrm{m}}}}\,\mathrm{d}\boldsymbol{\xi}$$

$$\mathrm{d}\boldsymbol{\xi} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$$

posterior



$$\mathbf{u}|\mathbf{x} = \mathcal{N}\left(\boldsymbol{\mu}(\mathbf{x})\,, \boldsymbol{\Sigma}\right)$$

$$\boldsymbol{\Sigma} = \left(\mathbf{C}^{-1} + \frac{1}{\sigma_{\mathrm{x}}^2}\,\mathbf{A}\,\mathbf{A}^{\mathsf{T}}\right)^{-1}$$

$$\boldsymbol{\mu}(\mathbf{x}) = \frac{1}{\sigma_{\mathrm{x}}^2}\,\boldsymbol{\Sigma}\,\mathbf{A}^{\mathsf{T}}\,\mathbf{x}$$

stationary distribution

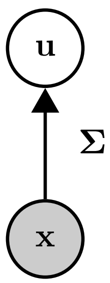
$$\mathbf{u}|\mathbf{x} = \mathcal{N}\left(\tilde{\boldsymbol{\mu}}(\mathbf{x})\,, \tilde{\boldsymbol{\Sigma}}\right)$$

$$(\mathbf{W} - \mathbf{I})\;\tilde{\boldsymbol{\Sigma}} + \tilde{\boldsymbol{\Sigma}}\;(\mathbf{W} - \mathbf{I}) = -2\sigma_{\xi}^2\,\mathbf{I}$$

$$\tilde{\boldsymbol{\mu}}(\mathbf{x}) = (\mathbf{I} - \mathbf{W})^{-1}\;\mathbf{F}\,\mathbf{x}$$

LANGEVIN SAMPLING IS ~~VERY~~ maximally SLOW

Hennequin et al, NIPS 2014

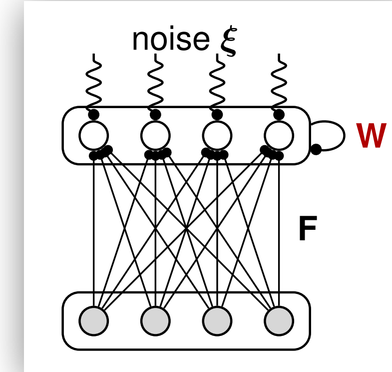


$$\mathbf{u}|\mathbf{x} = \mathcal{N}(\boldsymbol{\mu}(\mathbf{x}), \boldsymbol{\Sigma})$$

$$\boldsymbol{\Sigma} = \left(\mathbf{C}^{-1} + \frac{1}{\sigma_{\mathrm{x}}^2}\mathbf{A}\,\mathbf{A}^{\mathsf{T}}\right)^{-1}$$

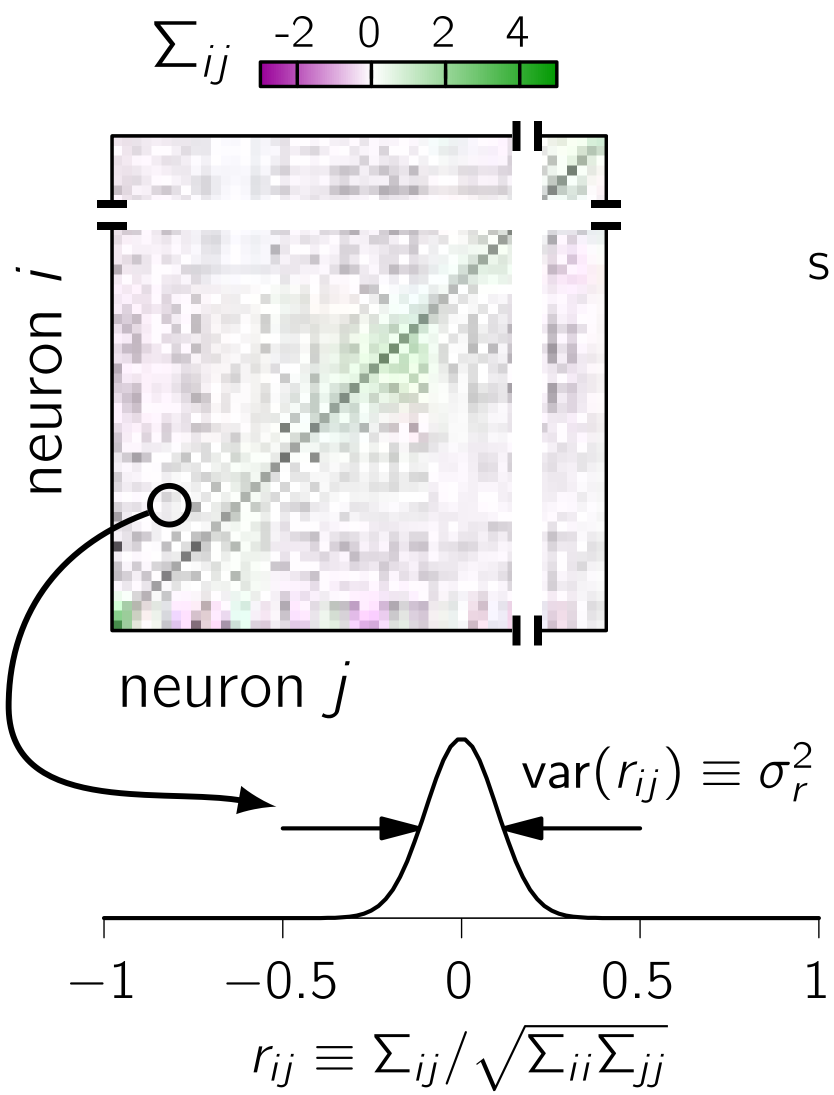
$$\boldsymbol{\mu}(\mathbf{x}) = \frac{1}{\sigma_{\mathrm{x}}^2}\boldsymbol{\Sigma}\,\mathbf{A}^{\mathsf{T}}\mathbf{x}$$

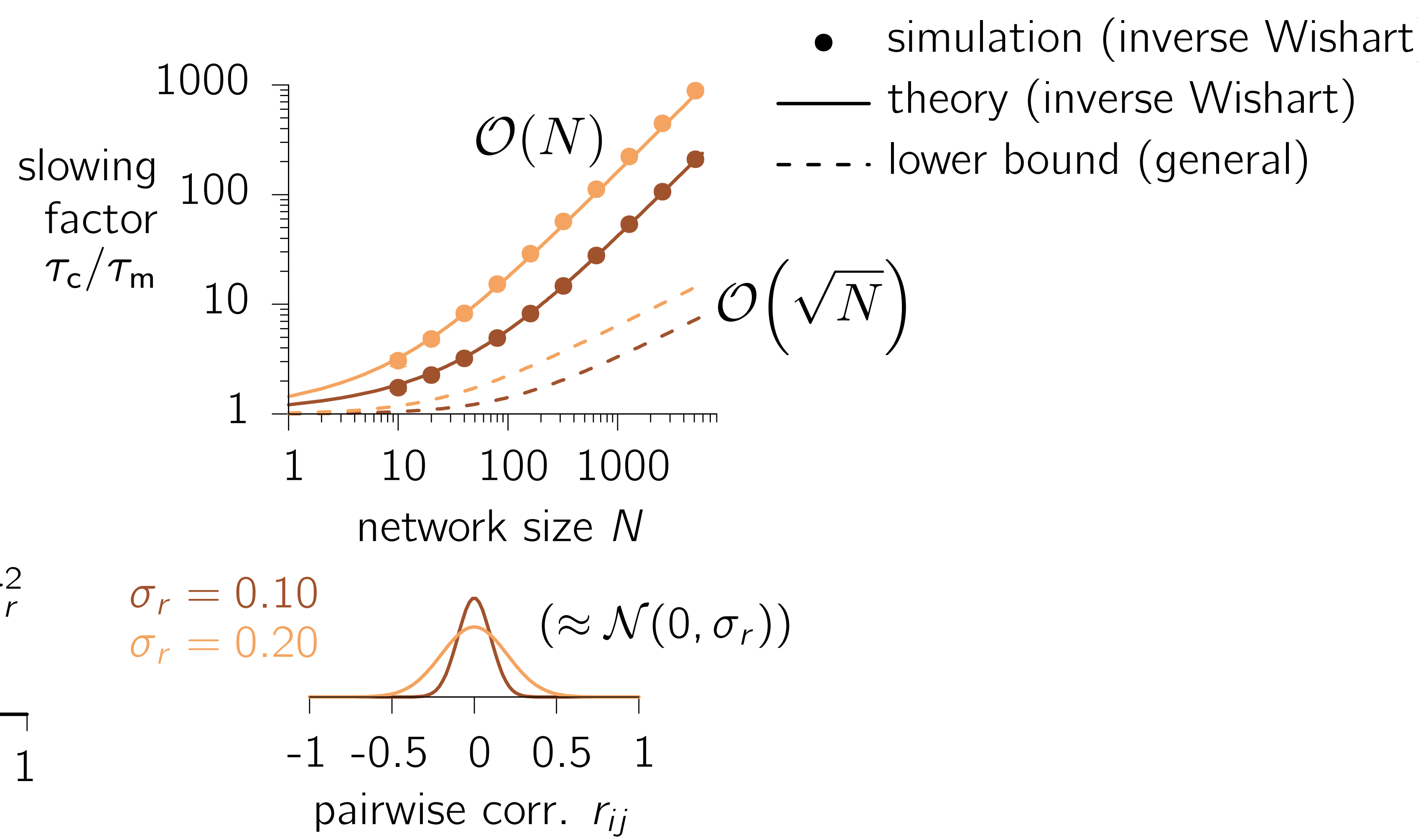
standard solution (Langevin, Gibbs):
symmetric weight matrix → detailed balance



$$\mathbf{W} = \mathbf{I} - \sigma_\xi^2\,\boldsymbol{\Sigma}^{-1} \qquad \text{and} \qquad \mathbf{F} = \frac{\sigma_\xi^2}{\sigma_{\mathrm{x}}^2}\mathbf{A}^{\mathsf{T}}$$

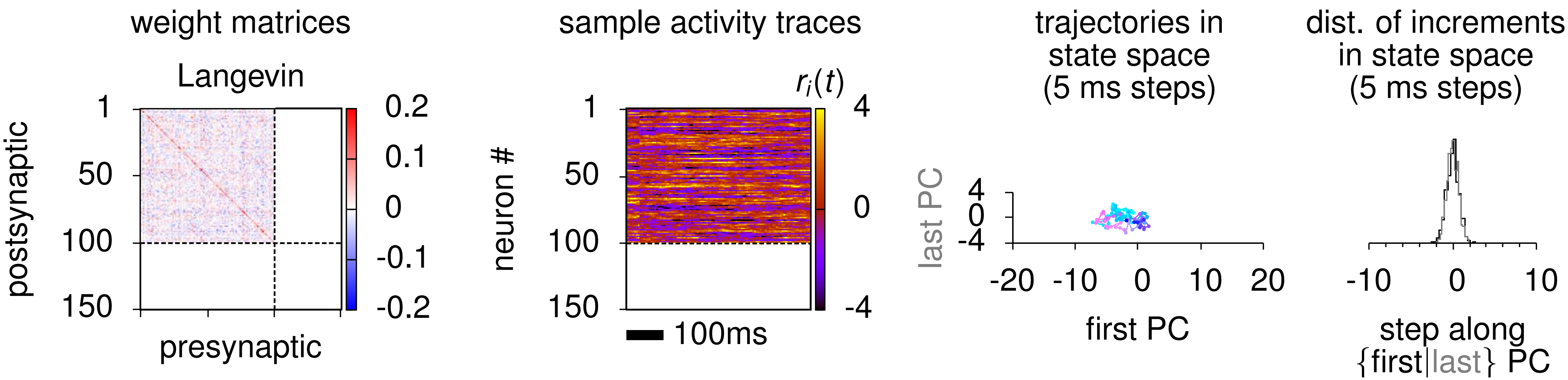
dynamics slow down horribly as N grows!





THE PROBLEM WITH LANGEVIN

Hennequin et al, NIPS 2014



A MORE GENERAL SOLUTION

Hennequin et al, NIPS 2014

$$\mathbf{W} = \mathbf{I} - \sigma_\xi^2\, \boldsymbol{\Sigma}^{-1} + \mathbf{S}\, \boldsymbol{\Sigma}^{-1} \qquad \forall \quad \mathbf{S}^\mathsf{T} = -\mathbf{S}$$

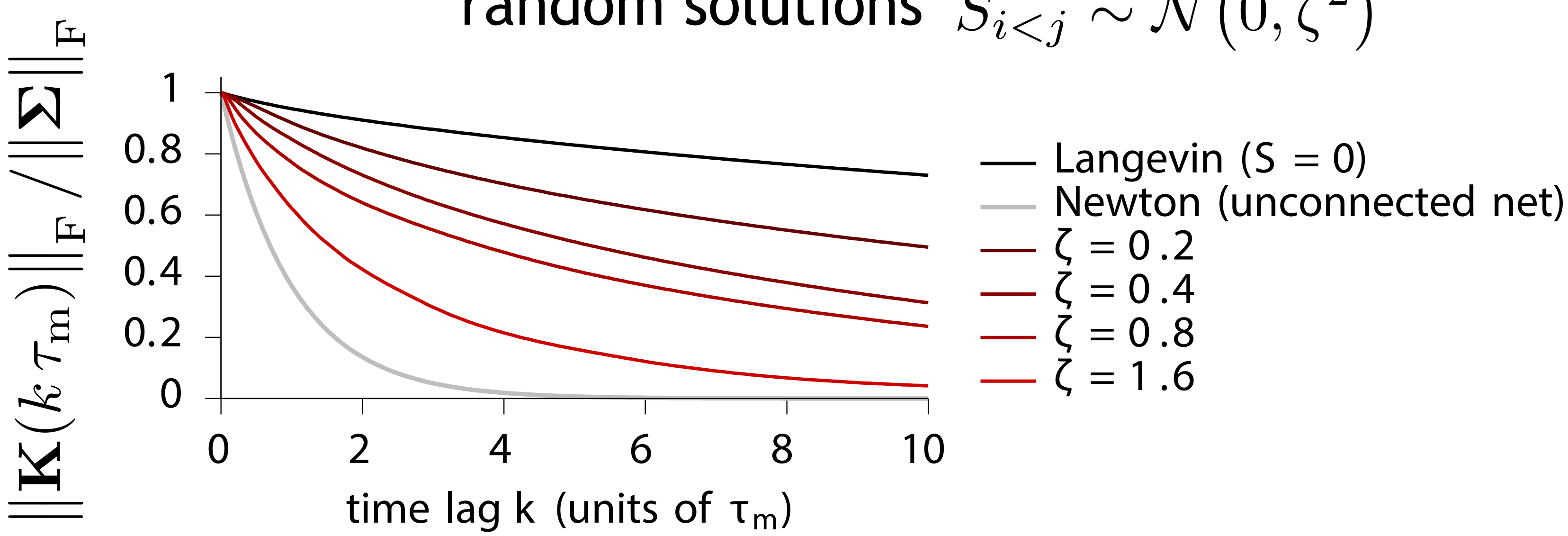
extra $\mathcal{O}(N^2)$ degrees of freedom!

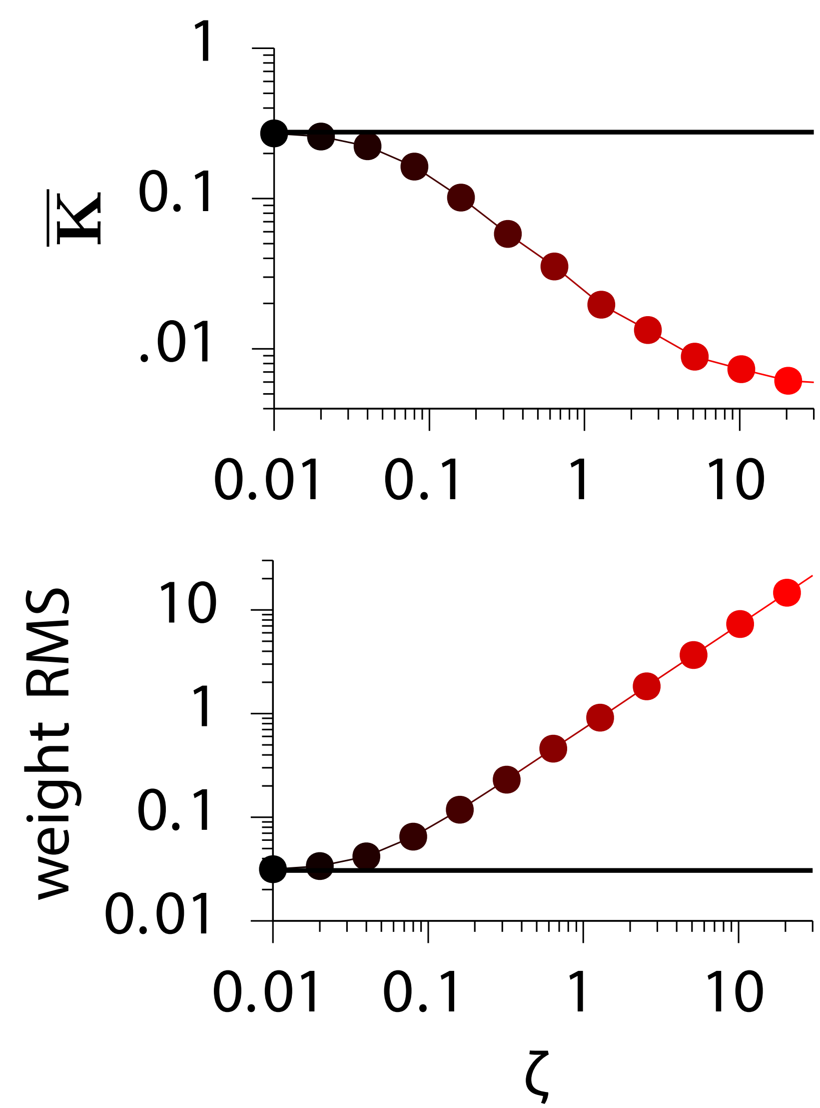
a measure of slowness: total autocovariance $\left\|\mathbf{K}(\tau)\right\|_\mathrm{F}$ $\longrightarrow$ **total slowness**

where $\mathbf{K}(\tau) = \left\langle \mathbf{x}(t+\tau)\ \mathbf{x}(t)^\mathsf{T} \right\rangle_t$ and $\left\|\mathbf{M}\right\|_\mathrm{F}^2 = \sum_{ij} \left|M_{ij}\right|^2$

$$\overline{\mathbf{K}} = \frac{\int_0^\infty \left\|\mathbf{K}(\tau)\right\|_\mathrm{F}^2\, \mathrm{d}\tau}{2\, N\, \tau_\mathrm{m}}$$

random solutions $S_{i<j} \sim \mathcal{N}\left(0, \zeta^2\right)$





OPTIMIZE FOR SPEED

Hennequin et al, NIPS 2014

$$\mathbf{W} = \mathbf{I} - \sigma_\xi^2\,\mathbf{\Sigma}^{-1} + \mathbf{S}\,\mathbf{\Sigma}^{-1} \qquad \forall \quad \mathbf{S}^{\mathsf{T}} = -\mathbf{S}$$

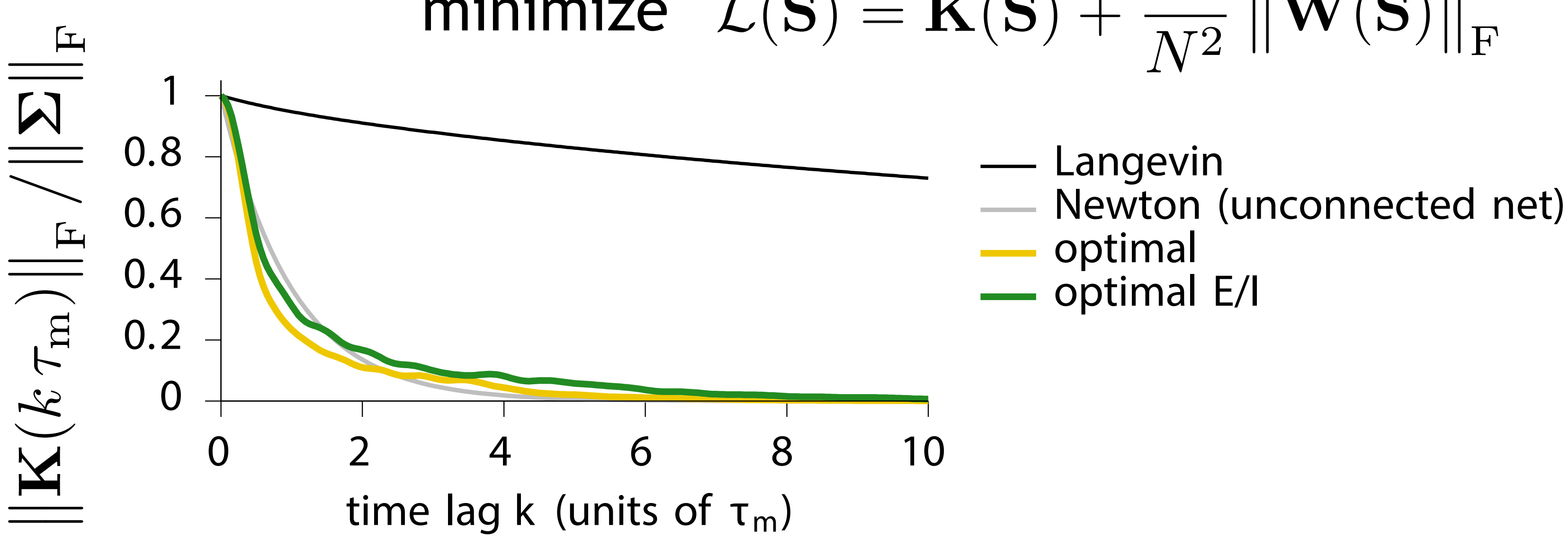
extra $\mathcal{O}(N^2)$ degrees of freedom!

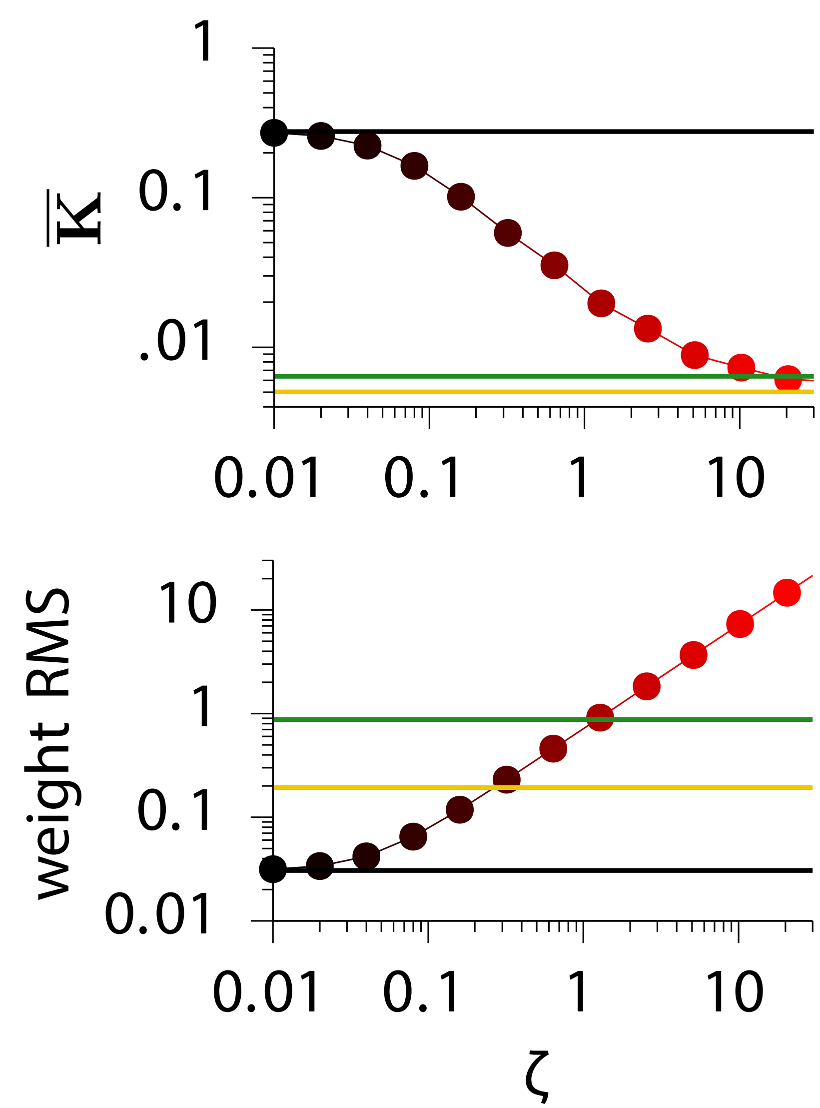
a measure of slowness: total autocovariance $\|\mathbf{K}(\tau)\|_{\mathrm{F}}$ $\longrightarrow$ **total slowness**

where $\mathbf{K}(\tau) = \left\langle \mathbf{x}(t+\tau)\;\mathbf{x}(t)^{\mathsf{T}} \right\rangle_t$ and $\|\mathbf{M}\|_{\mathrm{F}}^2 = \sum_{ij} |M_{ij}|^2$

$$\overline{\mathbf{K}} = \frac{\int_0^\infty \|\mathbf{K}(\tau)\|_{\mathrm{F}}^2\,\mathrm{d}\tau}{2\,N\,\tau_{\mathrm{m}}}$$

minimize $\mathcal{L}(\mathbf{S}) = \overline{\mathbf{K}}(\mathbf{S}) + \frac{\lambda}{N^2}\,\|\mathbf{W}(\mathbf{S})\|_{\mathrm{F}}^2$

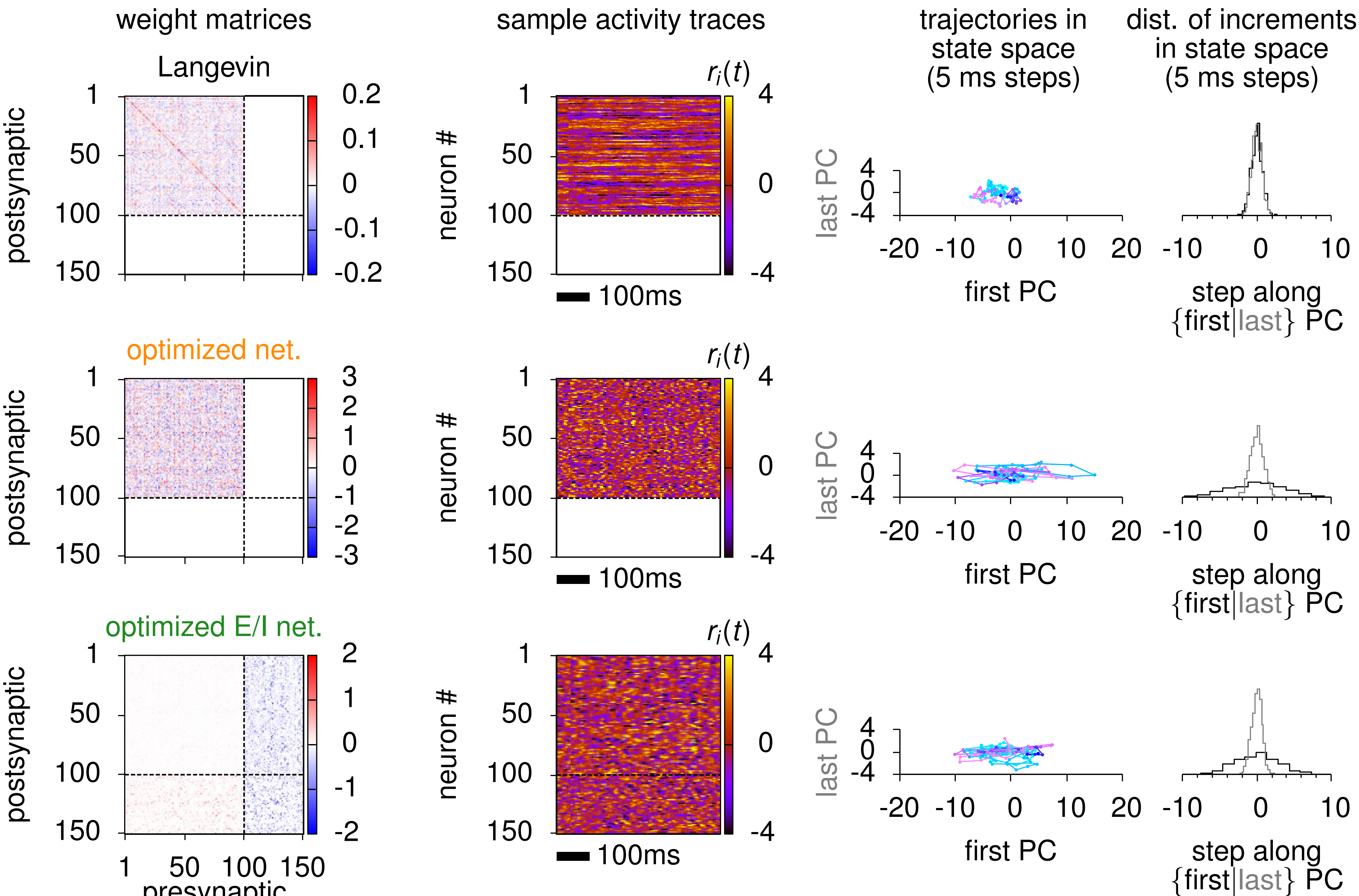




Murphy & Miller, 2009; Hennequin et al, 2014

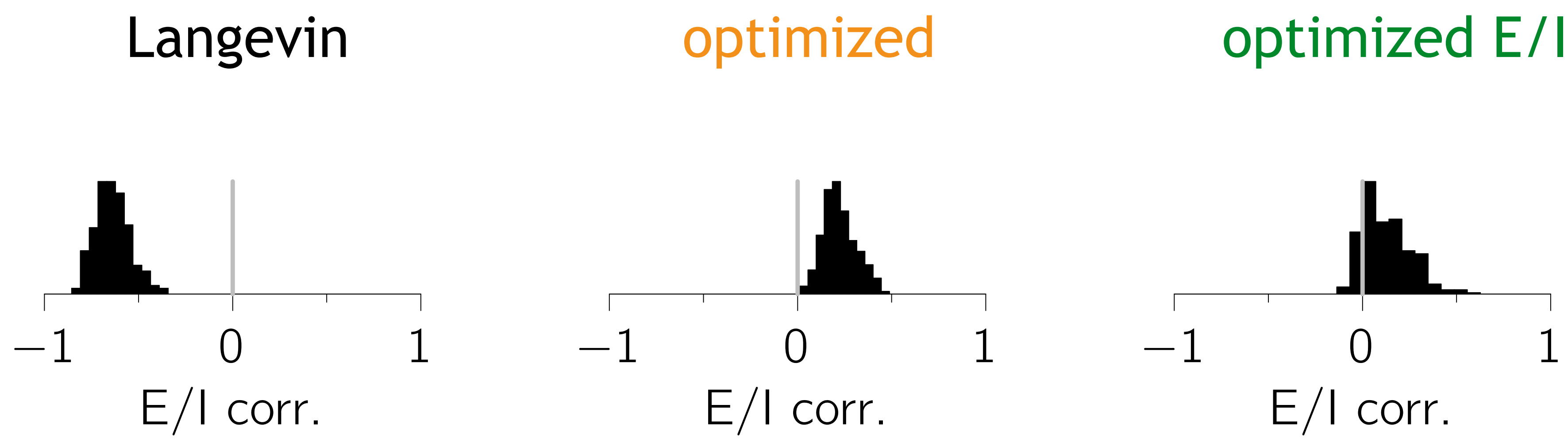
NON-NORMAL AMPLIFICATION TO THE RESCUE

Hennequin et al, NIPS 2014



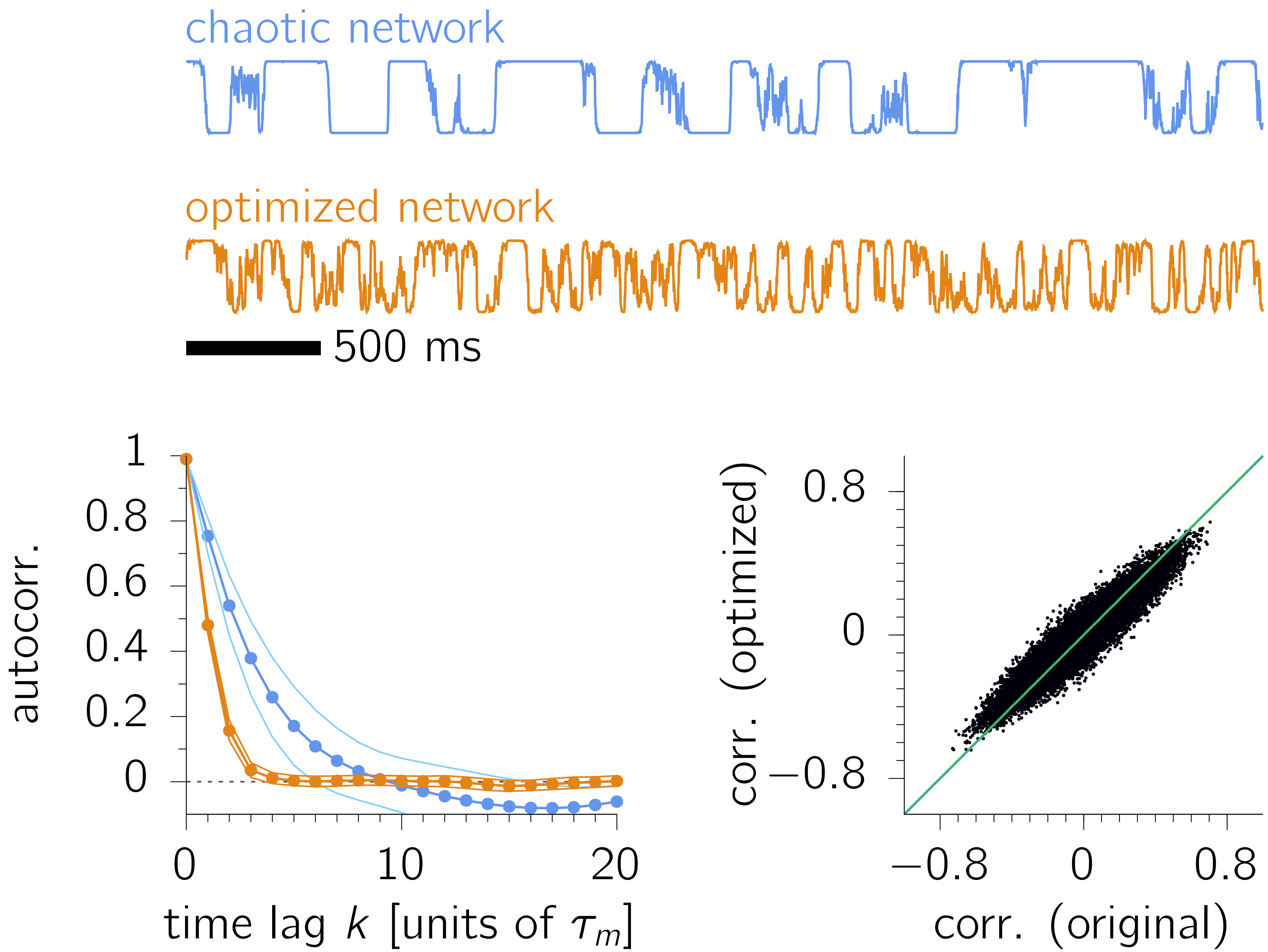
E/I BALANCE

Hennequin et al, NIPS 2014



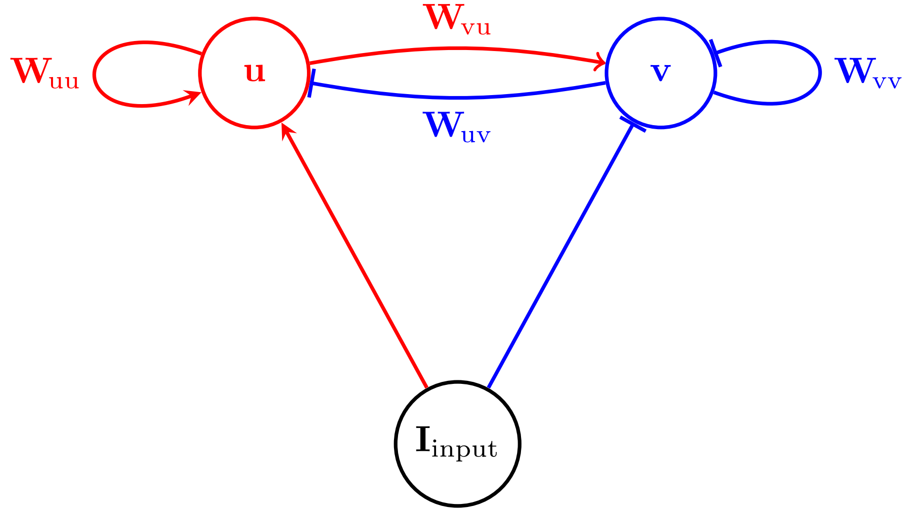
NONLINEAR CASE

Hennequin et al, NIPS 2014



E/I NETWORKS FOR SAMPLING

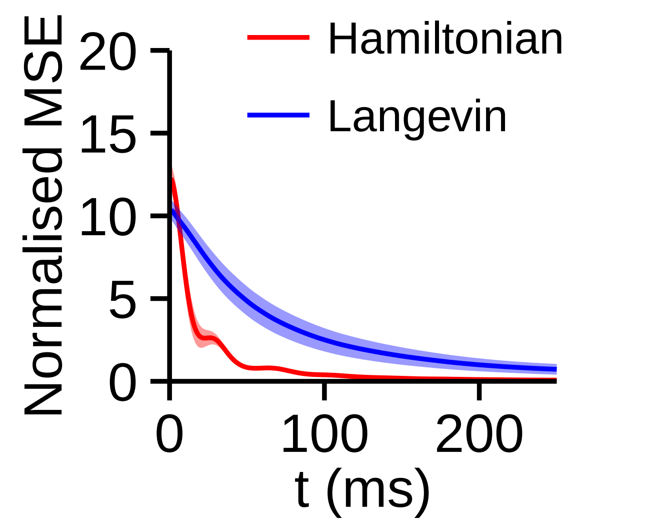
Aitchison & Lengyel, arXiv 2014, in prep



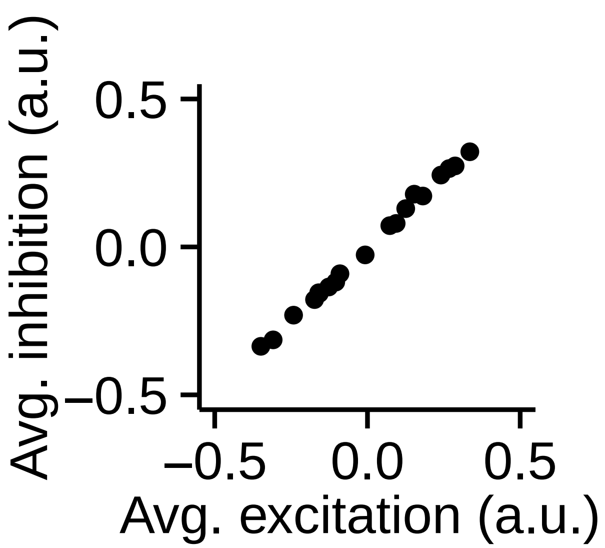
=

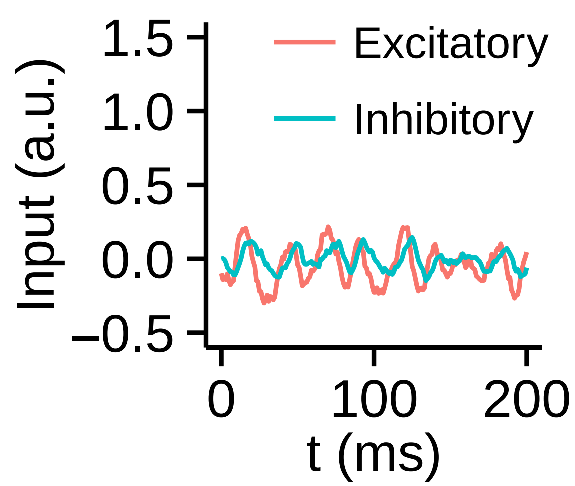
Hamiltonian Monte Carlo

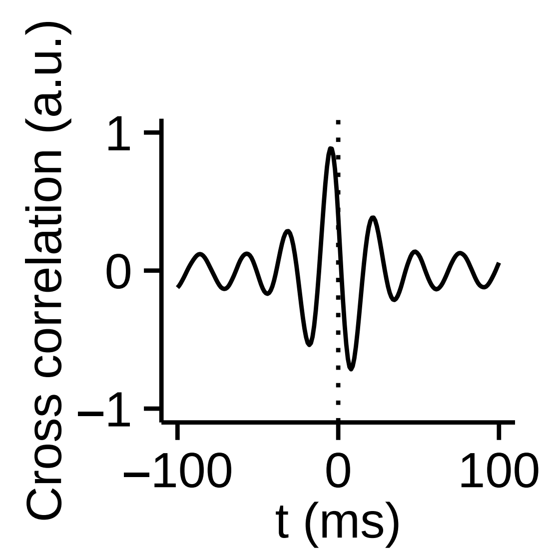
Duane et al 1987, Neal 2011



E/I BALANCE AND OSCILLATIONS



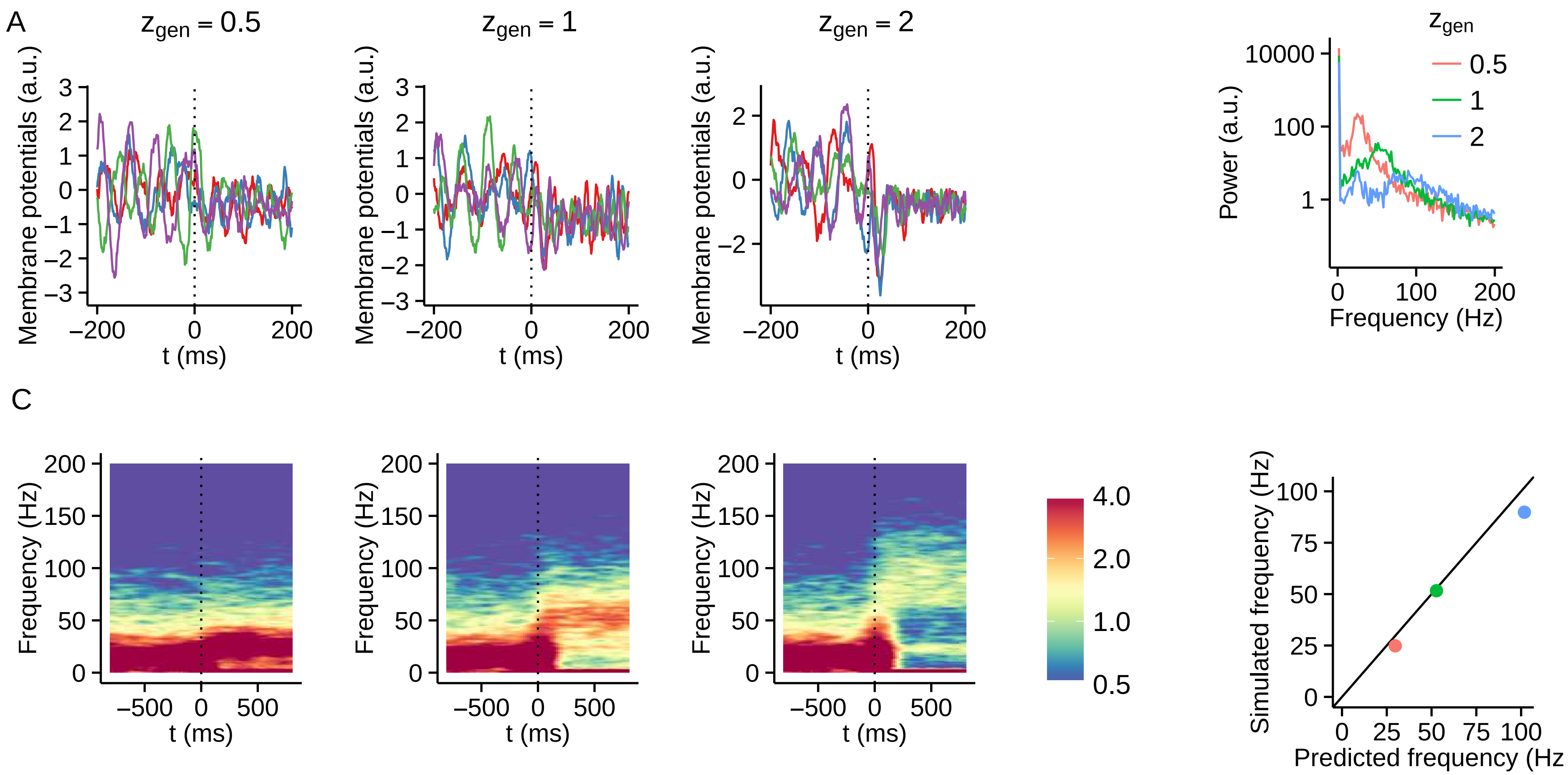




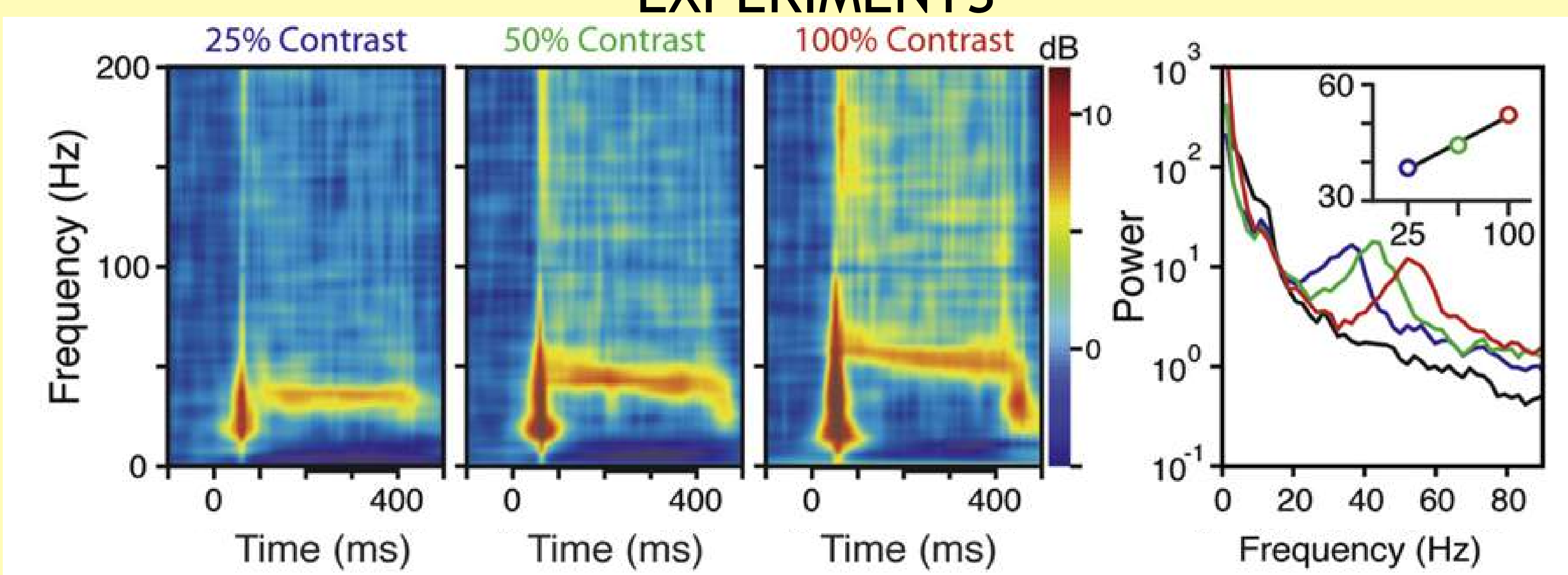
E/I NETWORKS FOR SAMPLING (HAMILTONIAN)

Aitchison & Lengyel, arXiv 2014, in prep

OSCILLATION FREQUENCY DEPENDS ON CONTRAST



EXPERIMENTS

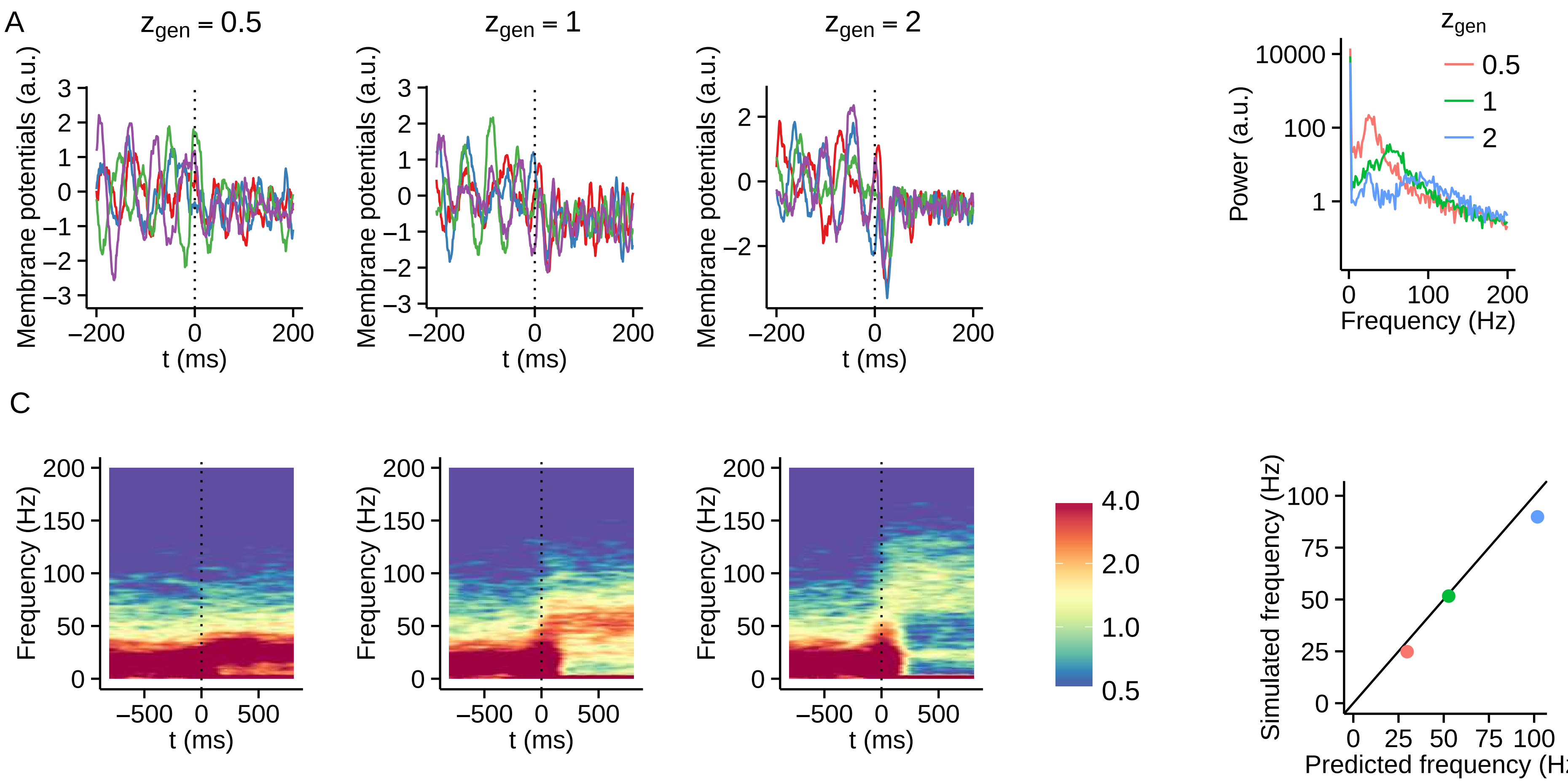


Ray & Maunsell, Neuron 2010

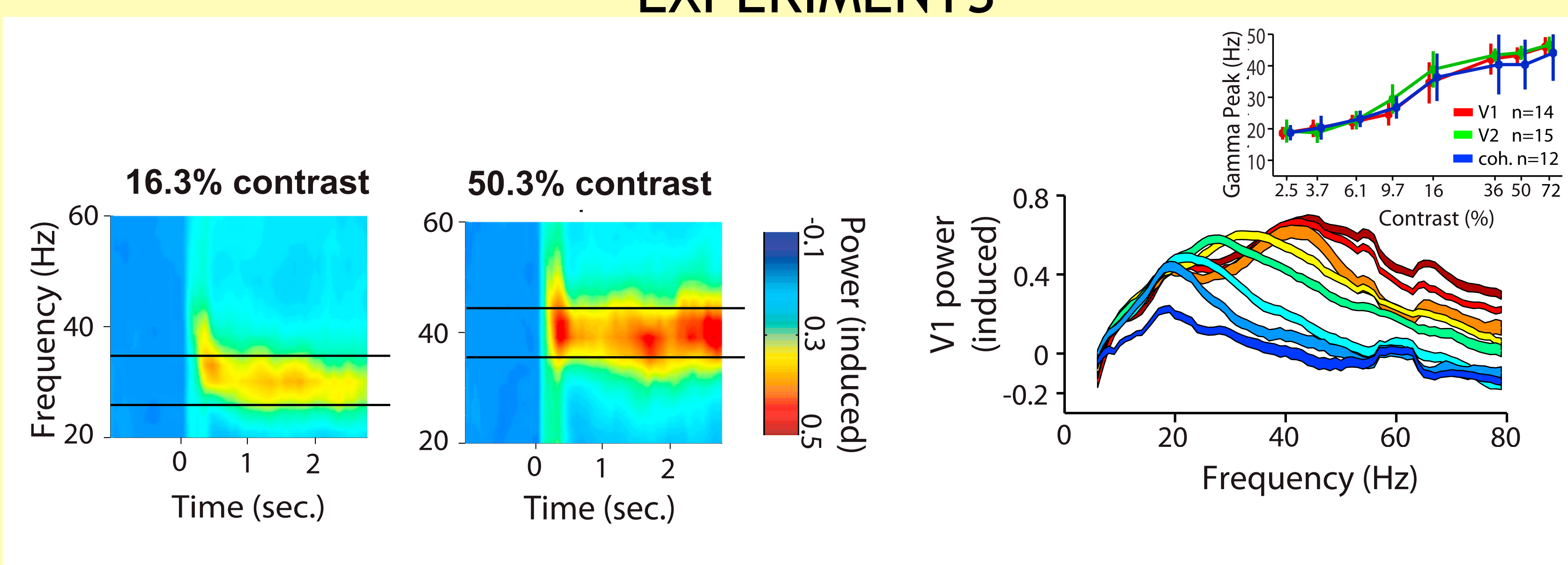
E/I NETWORKS FOR SAMPLING (HAMILTONIAN)

Aitchison & Lengyel, arXiv 2014, in prep

OSCILLATION FREQUENCY DEPENDS ON CONTRAST



EXPERIMENTS

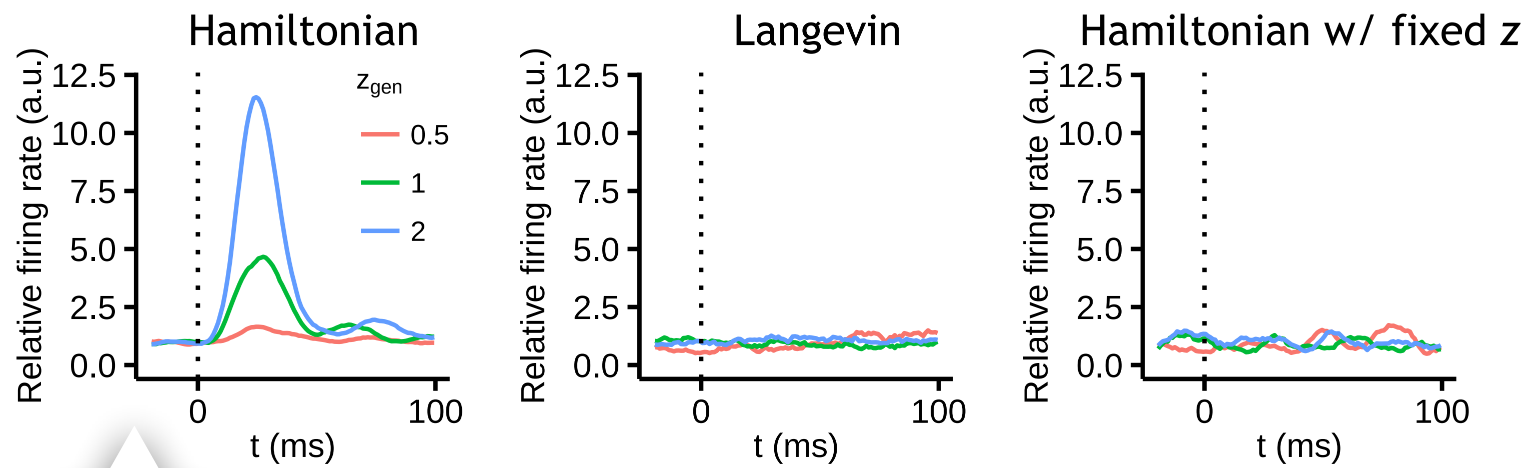


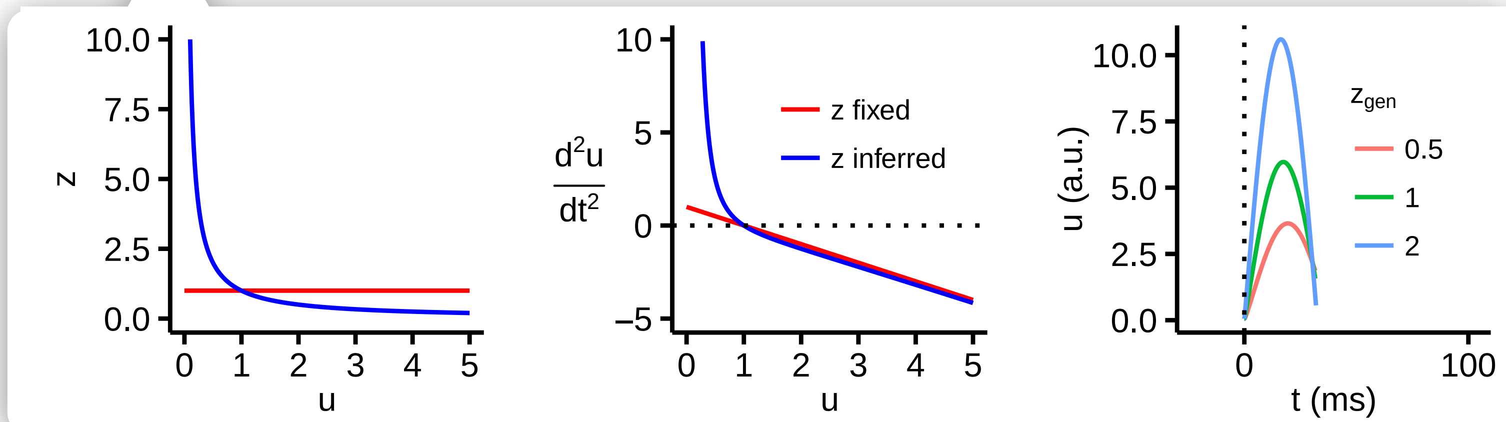
Roberts et al, Neuron 2013

E/I NETWORKS FOR SAMPLING (HAMILTONIAN)

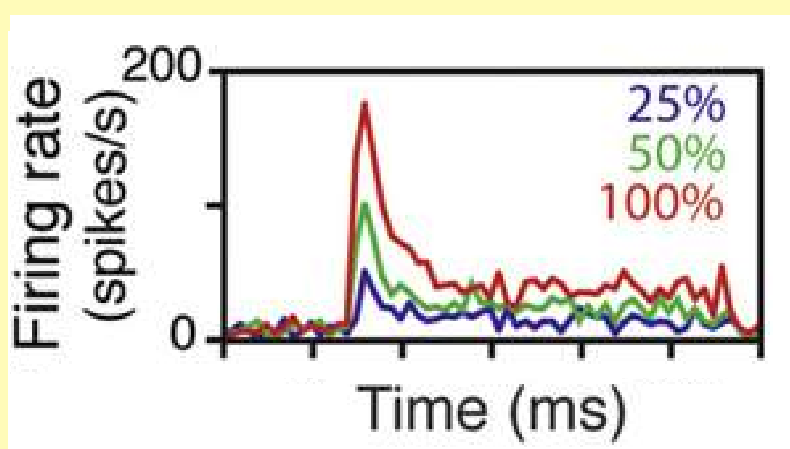
Aitchison & Lengyel, arXiv 2014, in prep

FIRING RATE TRANSIENTS





EXPERIMENTS



Ray & Maunsell, Neuron 2010

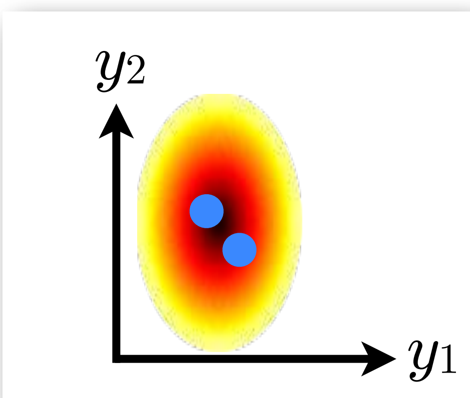
SLOW OSCILLATIONS

‘tempered transitions’

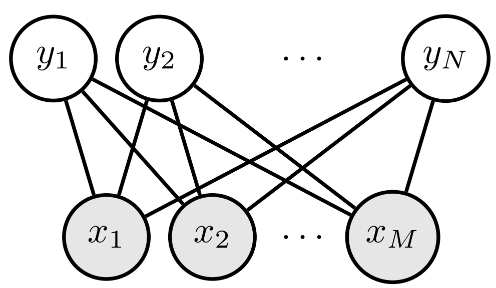
Neal 1996

$$\mathbf{r} \sim \mathrm{P}(\mathbf{y}=\mathbf{r}|\mathbf{x}) \simeq \left[\prod_i \mathrm{P}(y_i = r_i|\mathbf{x})\right] \left[\frac{\mathrm{P}(\mathbf{y}=\mathbf{r}|\mathbf{x})}{\prod_i \mathrm{P}(y_i = r_i|\mathbf{x})}\right]^{\alpha}$$

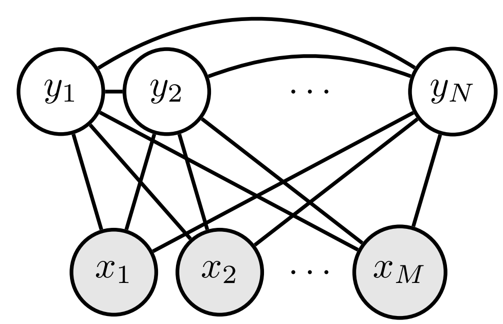
marginals (no correlations) + correlations



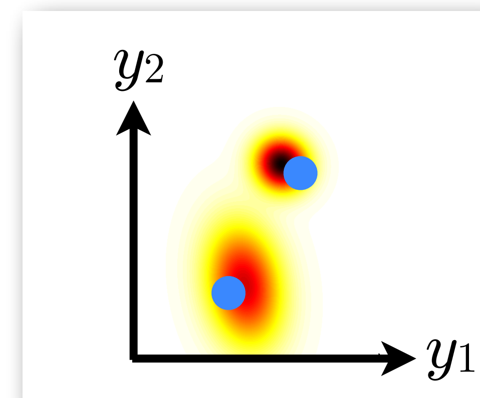
0 ⟵ α ⟶ 1



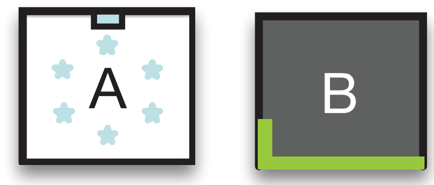
feed-forward



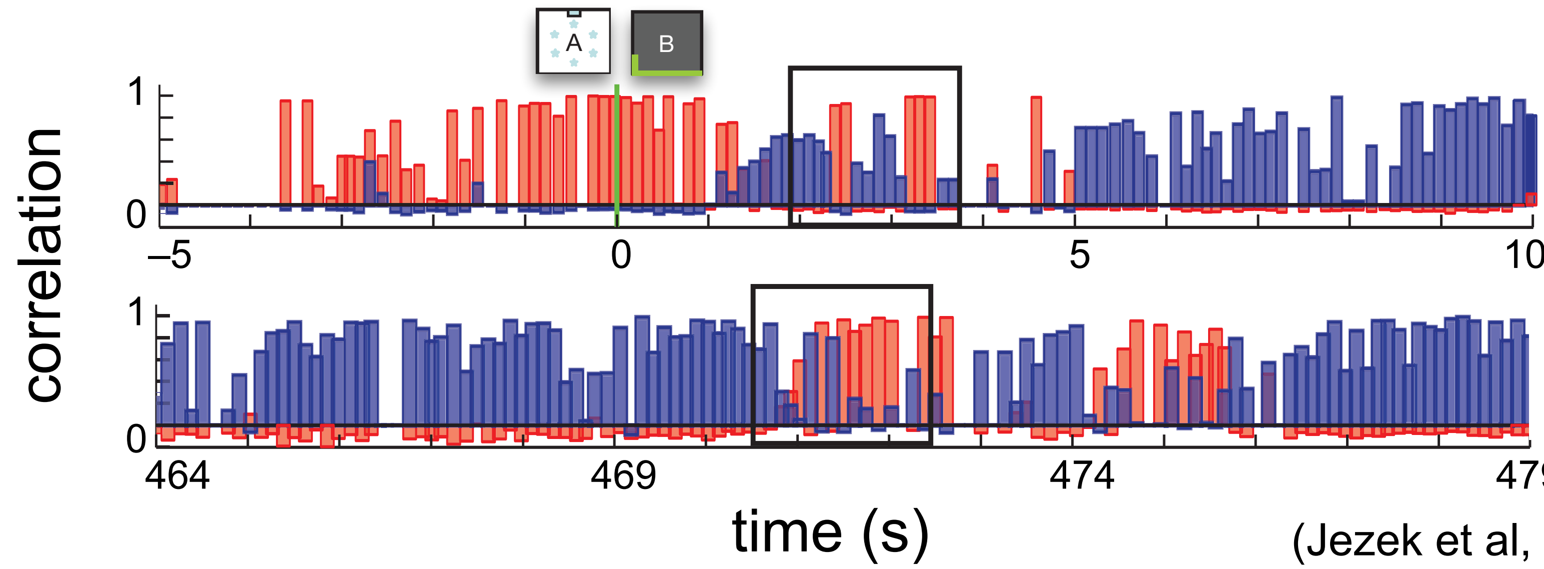
+ recurrents



HIPPOCAMPAL FLICKERING

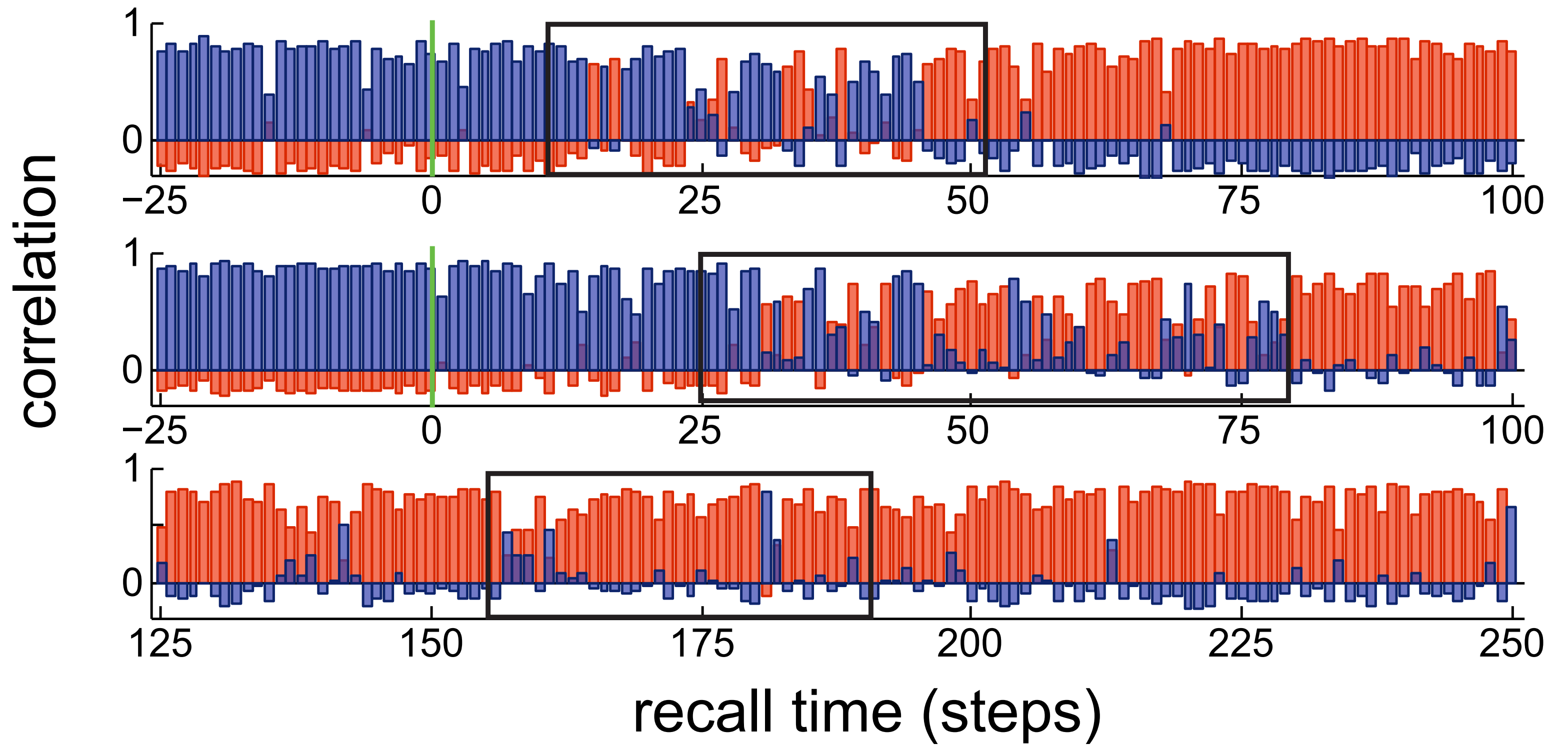


EXPERIMENT



(Jezek et al, 2011)

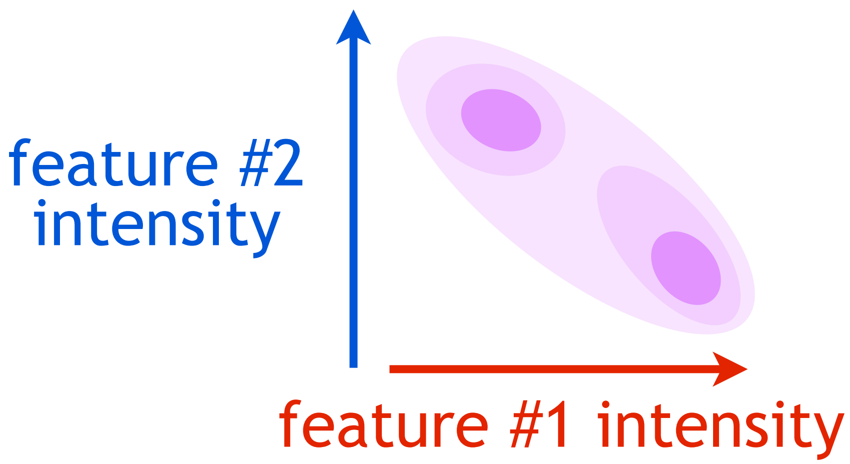
MODEL

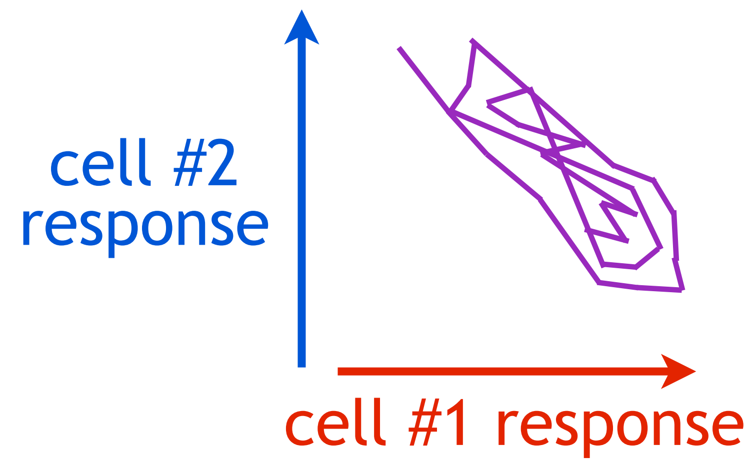


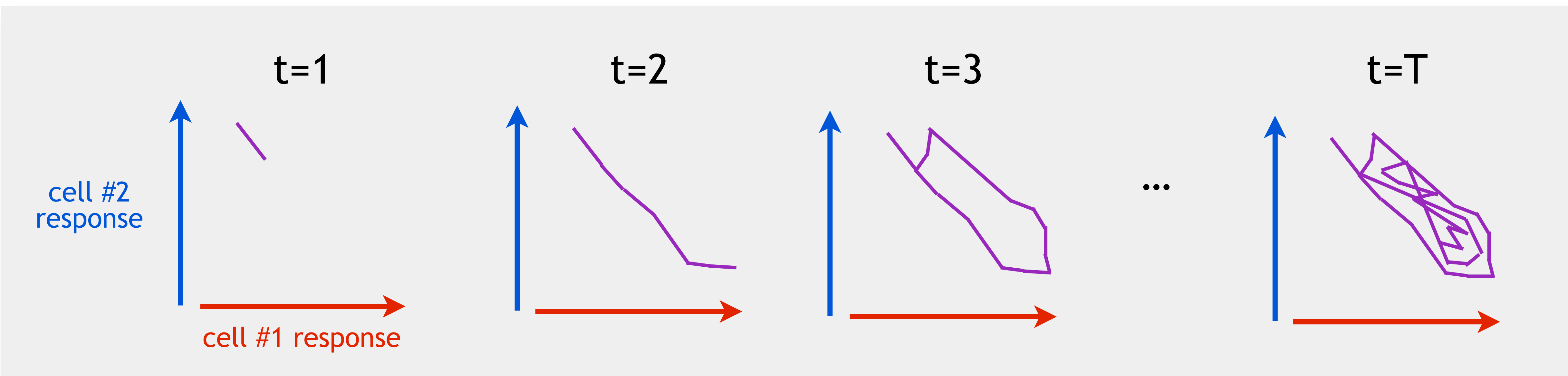
Savin et al, PLoS Comput Biol 2014

A PSYCHOPHYSICAL HALLMARK OF SAMPLING

Lengyel et al, arXiv 2015



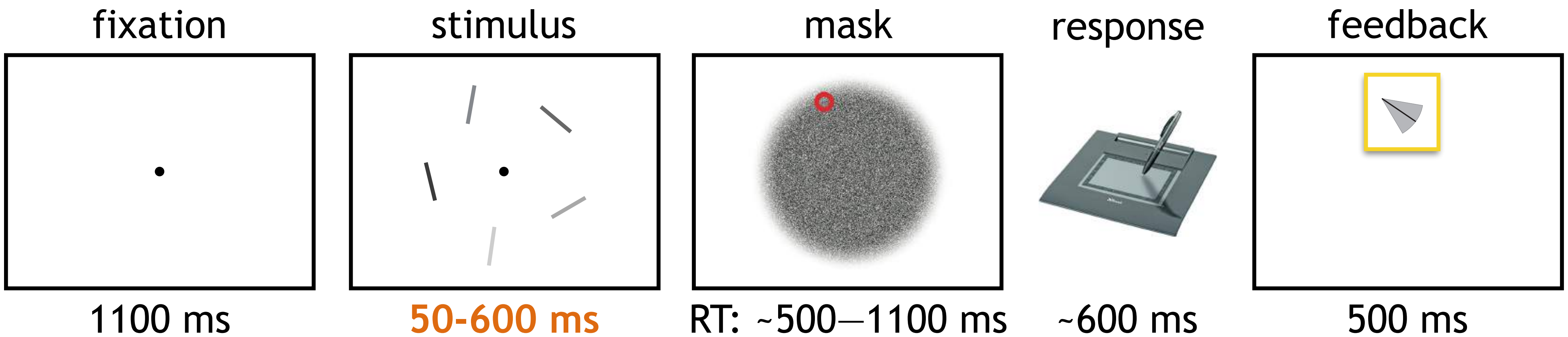


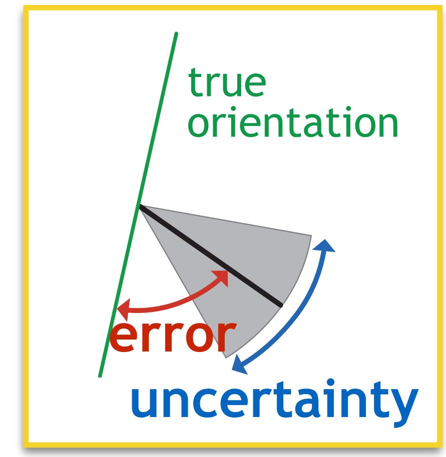


A GRADUAL REFINEMENT OF THE REPRESENTATION OF UNCERTAINTY

A PSYCHOPHYSICAL TEST

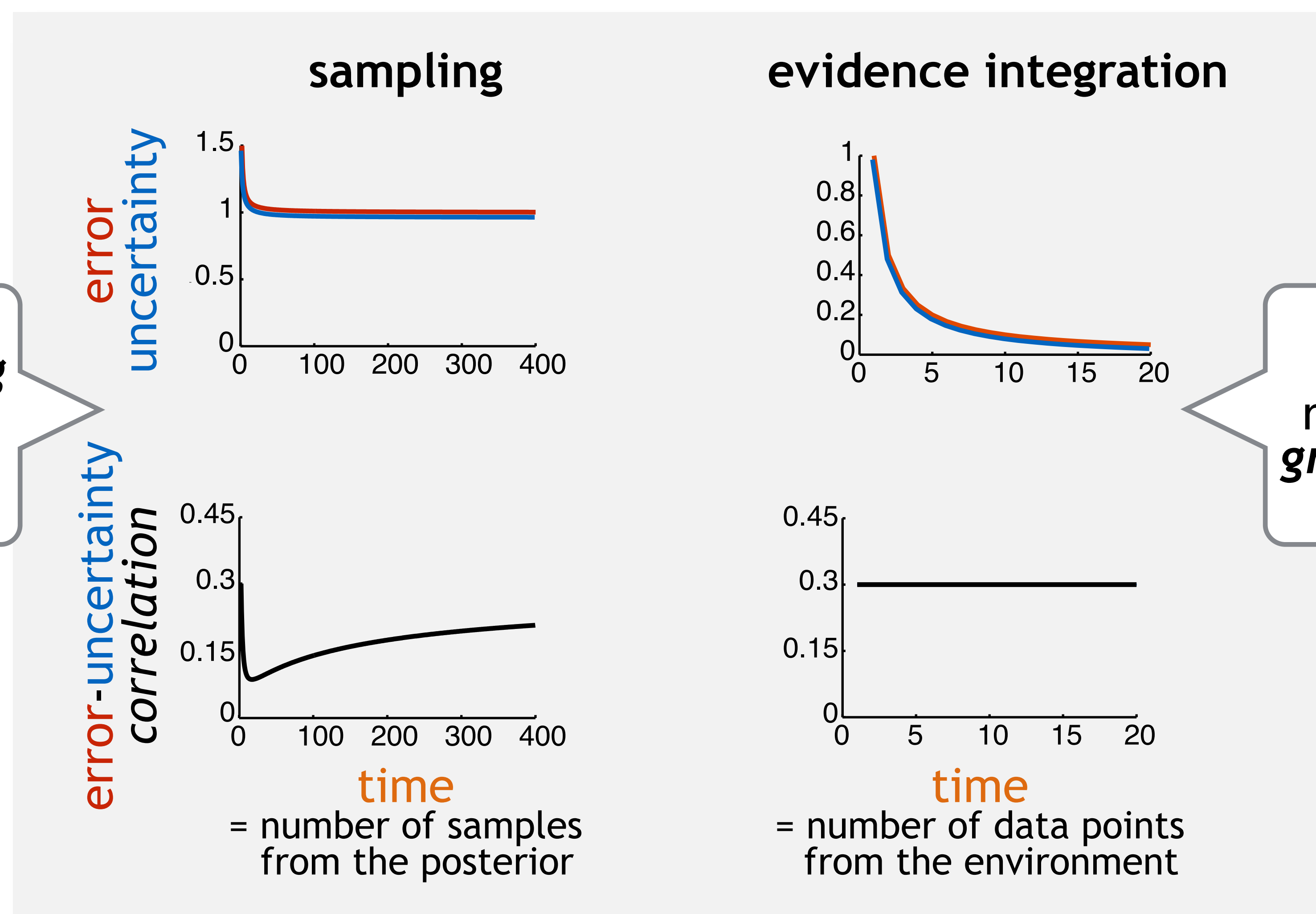
Lengyel et al, arXiv 2015





quality of information available: error (uncertainty)

quality of probabilistic representation: error-uncertainty *correlation*



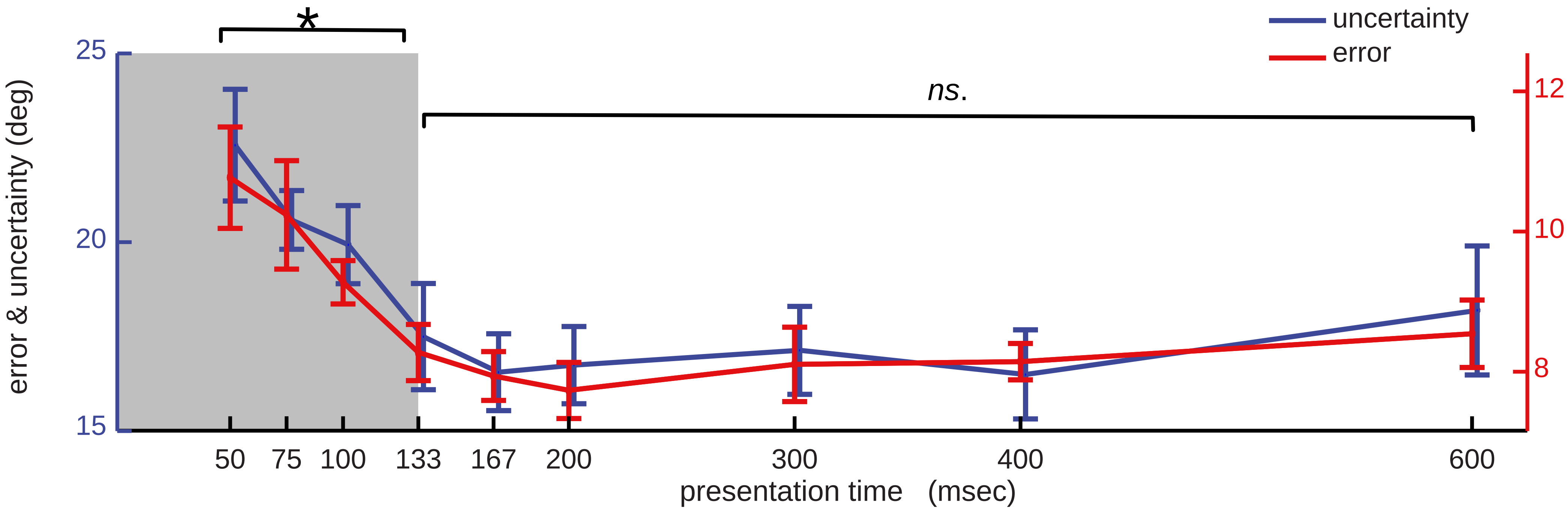
a ***gradually improving*** representation of a ***static*** posterior

a ***perfect*** representation of a ***gradually improving*** posterior

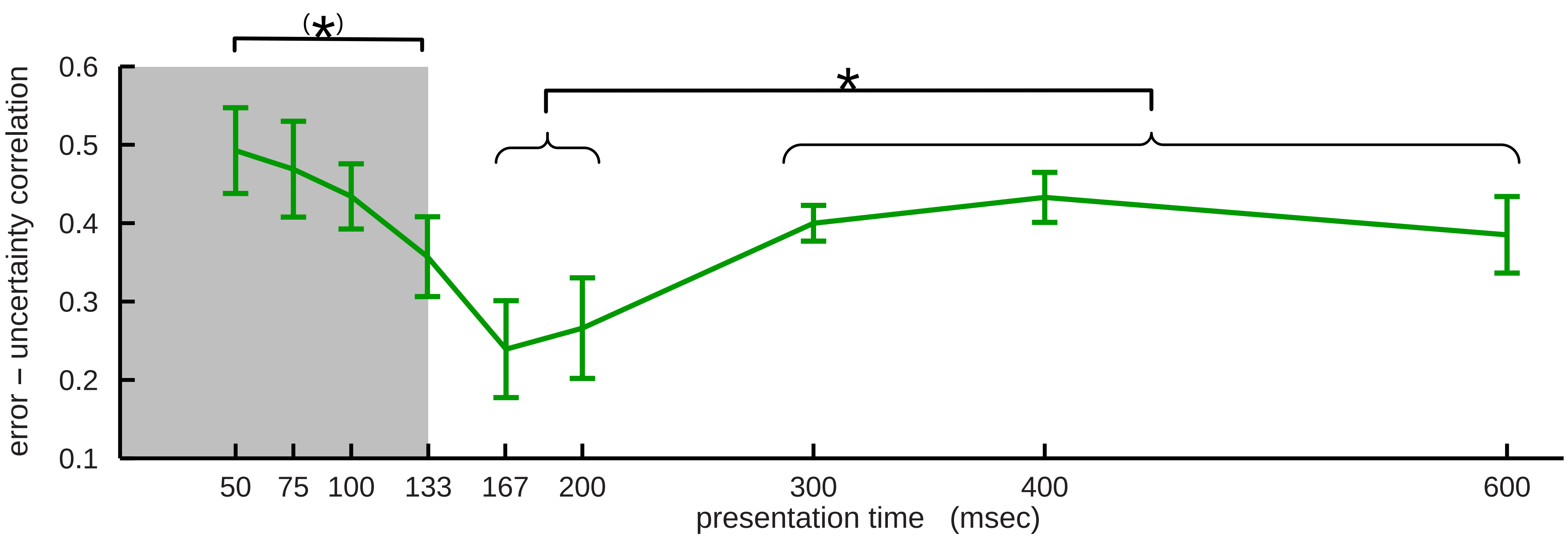
EFFECTS OF TIME

Lengyel et al, arXiv 2015

error and uncertainty



error-uncertainty correlation



SUMMARY

sampling

- is a simple and powerful way of representing uncertainty
- can be efficiently implemented by E/I neural circuit dynamics
- provides a natural account of
  - neural variability
  - the match between evoked and spontaneous activity
  - hippocampal flickering? → *Savin et al, PLoS Comput Biol 2014*

a new paradigm to obtain trial-by-trial measure of uncertainty:
humans' representation of uncertainty

- is well calibrated, multidimensional & uses a unitary scale
- reflects hallmarks of sampling (2-3 ms / sample)

ACKNOWLEDGEMENTS

Guillaume Hennequin

Cristina Savin

UCL

Laurence Aitchison

CEU

József Fiser

Ádám Koblinger

Marjena Popović

wellcome trust Investigator