Standard Models Fail to Capture Positional Invariances in Marmoset Foveal V1

REDWOOD CENTER for Theoretical Neuroscience

UC Berkeley Herbert Wertheim School of Optometry & Vision Science

Tejasvi Kothapalli[1,2], Ryan A. Ressmeyer[3], Declan Rowley[4], Jude F. Mitchell[5], Jacob L. Yates[1,2]

[1]Redwood Center for Theoretical Neuroscience, UC Berkeley, [2]Herbert Wertheim School of Optometry & Vision Science, UC Berkeley, [3]Department of Bioengineering, University of Washington, [4]Department of Psychiatry & Biobehavioral Sciences, UCLA, [5]University of Rochester

UC Berkeley

Introduction

Vision is actively sampled via eye movements. Even during fixations, the eyes continue to move, yet we are perceptually unaware of such movements [1]. One possibility is that fixational eye movements have limited effect on visual neurons with receptive fields (RFs) larger than the eye movements -- this is a core assumption of fixation tasks used widely in neuroscience. In contrast, phase sensitive neurons could be sensitive to spatial shifts much smaller than the receptive field throughout the visual system.

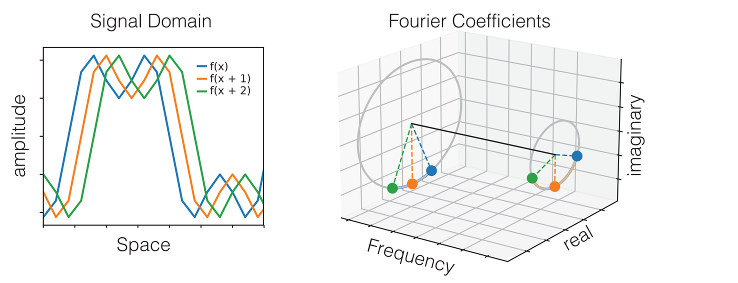
Goals

Quantify the sensitivity of central V1 neurons to fixational eye movements (FEMs)

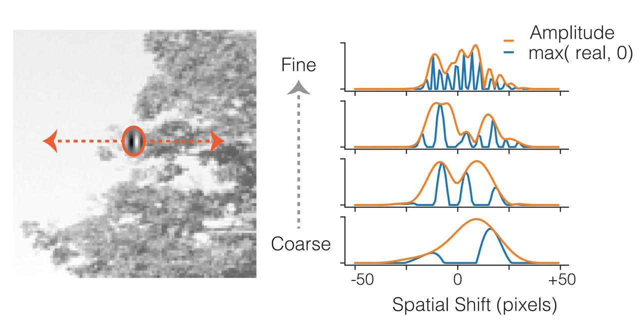
Test whether sensitivity is explained by standard mechanisms (RF size, Spatial Frequency Tuning, and Phase sensitivity)

Fixational eye movements as phase modulation

Local translation of a signal is equal to frequency-dependent changes in phase



Amplitude changes slowly across space in natural images



Complex cells are well-positioned to build invariance to FEMs on natural images by virtue of their invariance to spatial phase [2]

Approach

Behavior

- Subjects: 2 marmoset monkeys
- Fixation and Freeviewing Conditions
- High-resolution eye tracking: digital dual purkinje system [2, 3]

Electrophysiology

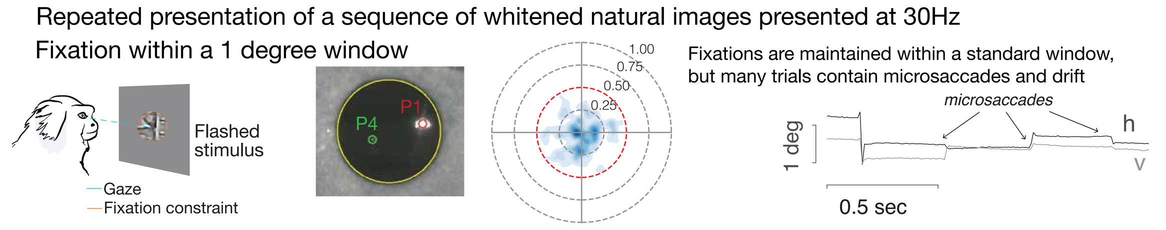
2 x 32 channel Neuronexus silicon arrays in foveal V1:
- spike sorting via Kilosort 4
- 10-20 single units per session. 80-120 reasonable units
- 20 sessions, 2400 units, 700 single units

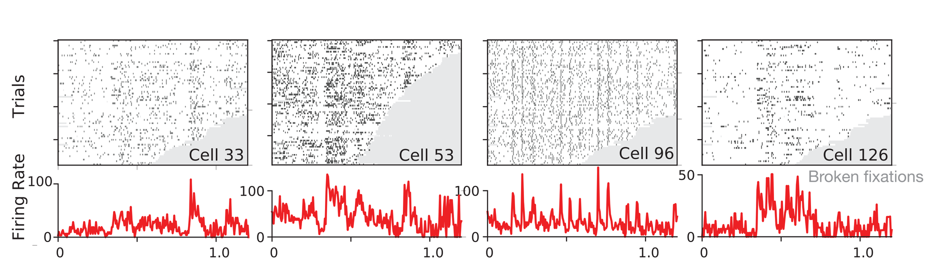
(Eyetracker available here) OpenIris DPI

Fixation Condition

Repeated presentation of a sequence of whitened natural images presented at 30Hz

Fixation within a 1 degree window



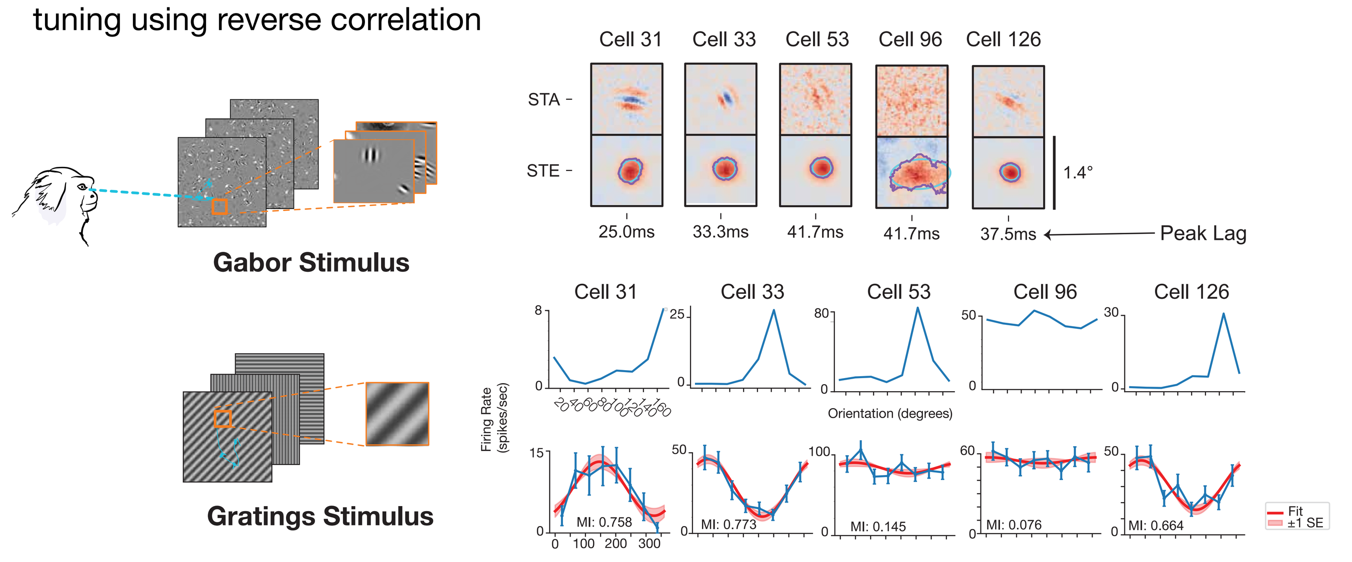


Unit Inclusion Criterion

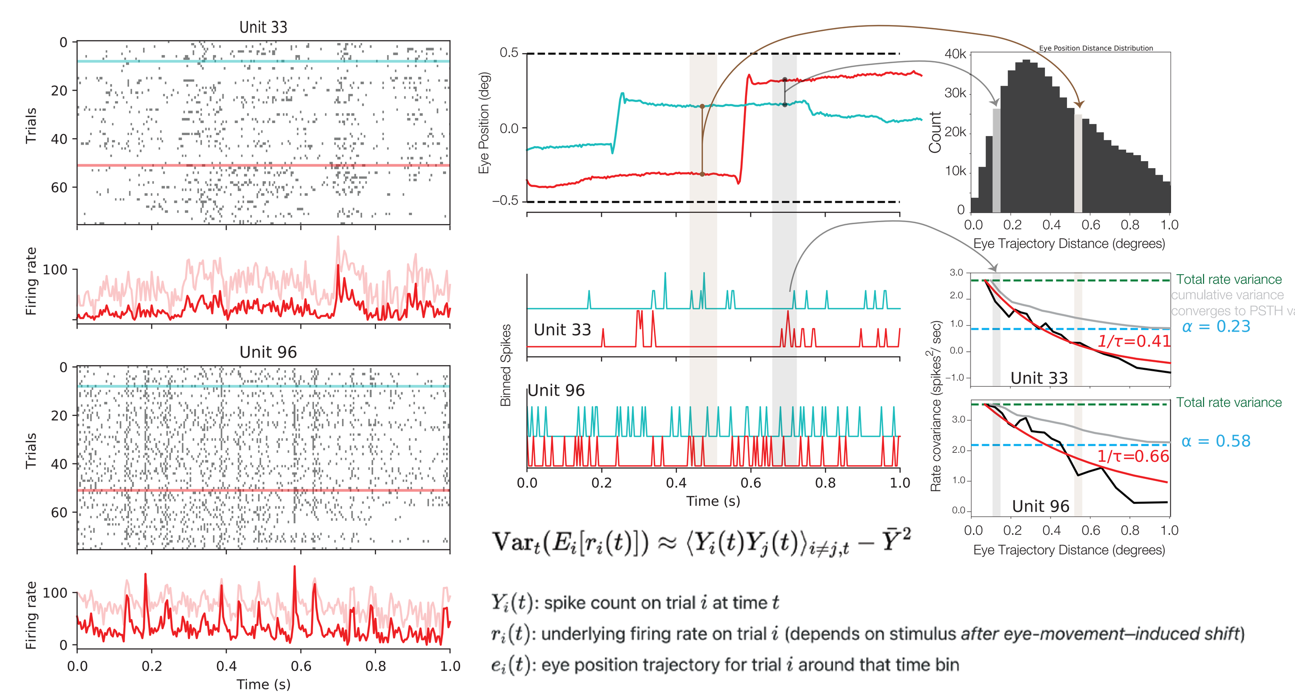
- Visually Responsive: STE SNR > 0.6
- More than 2.5 spikes/second in fixation condition
- Less than 60% contamination of spike refractory period

Characterizing Selectivity

Gaze-contingent region of interest (ROI) used to characterize spatial selectivity and tuning using reverse correlation



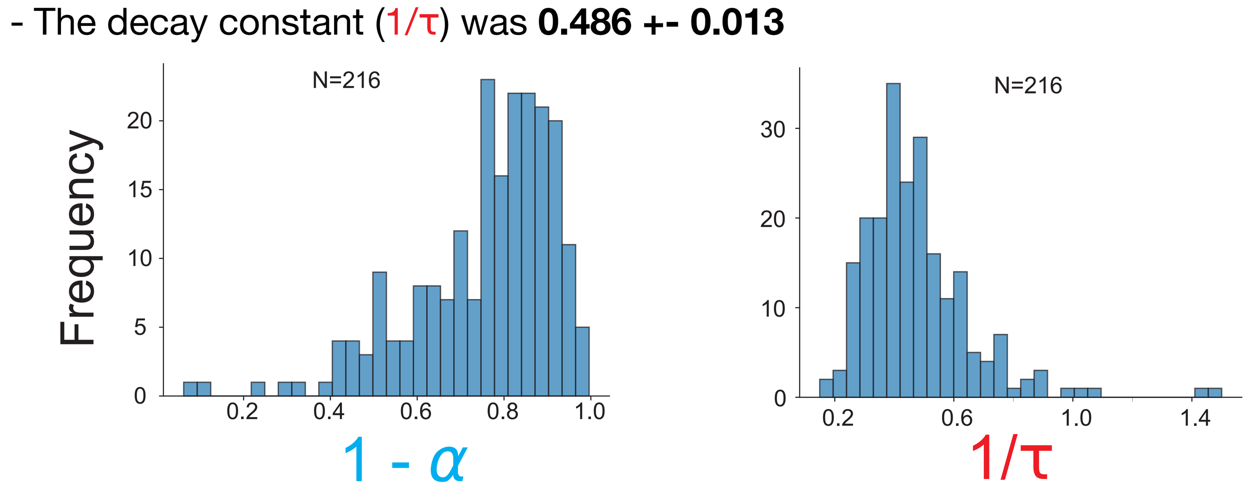
Measuring sensitivity to eye movements



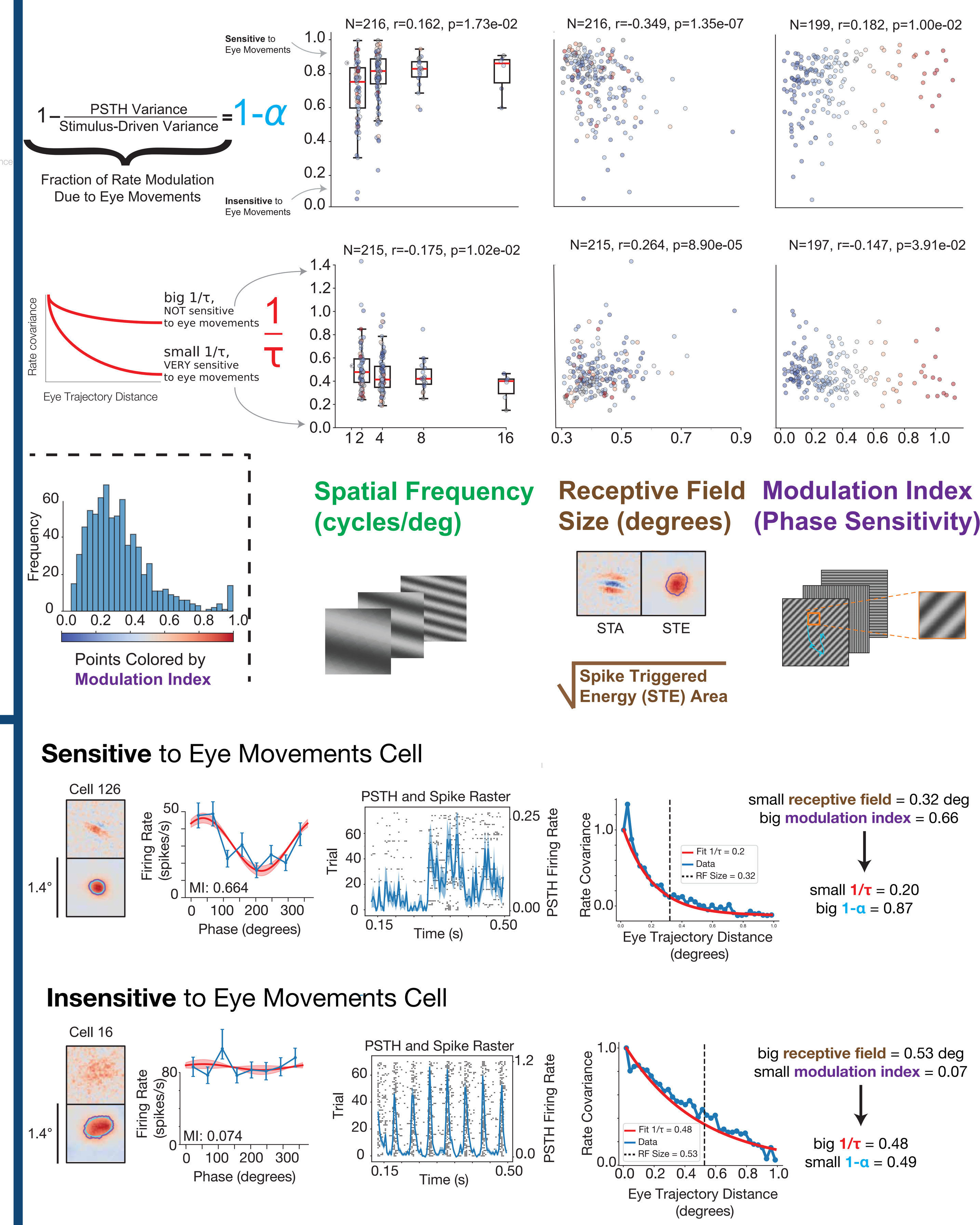
- Following analysis in [2], we used the Law of Total Variance to factorize the total variance for each cell into an eye-trajectory conditioned rate modulation and the modulation from the PSTH
- α is the fraction of stimulus-driven variance captured by the PSTH
- 1-α is the fraction of rate modulation due to eye movements
- 1/τ captures the dependence of rate modulation on eye trajectory distance

Eye movements drive a large fraction of the rate modulation

- Across the population, **75.6% +- 1.1%** of the rate modulation (1-α) is attributable to eye trajectory differences.
- The decay constant (1/τ) was **0.486 +- 0.013**

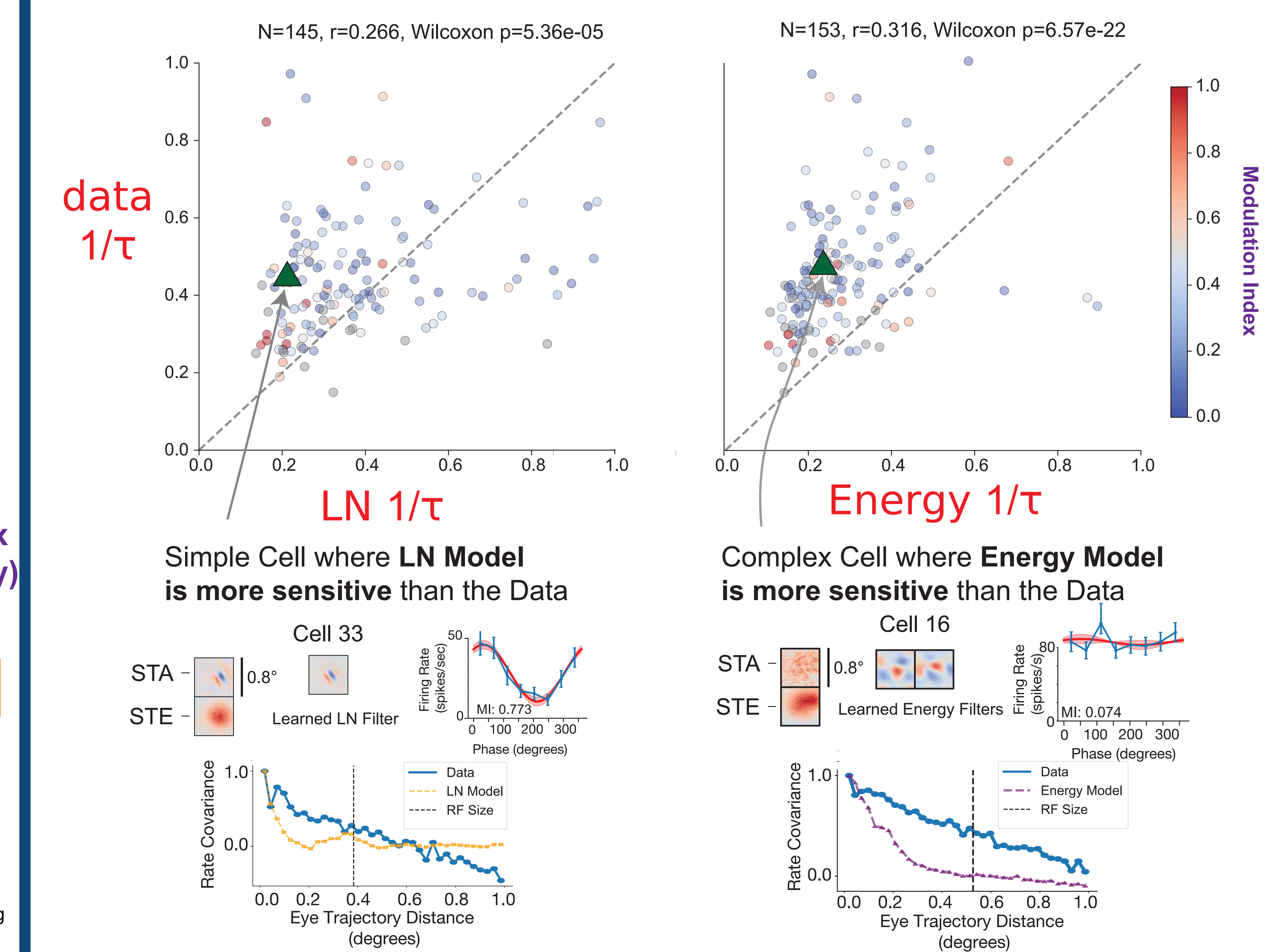


Eye movement sensitivity depends on receptive field properties



Standard models are more sensitive to eye movements than the real neurons

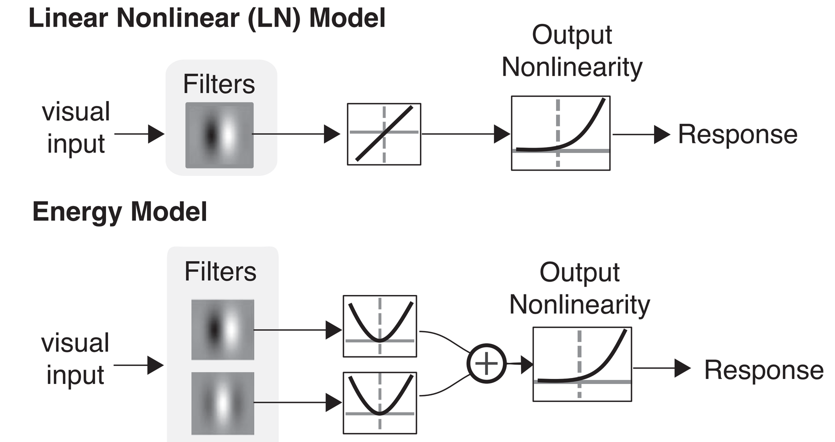
We use the learned LN and Energy Models' predicted spikes on the fixation condition to measure the models' eye movement sensitivity



Conclusions

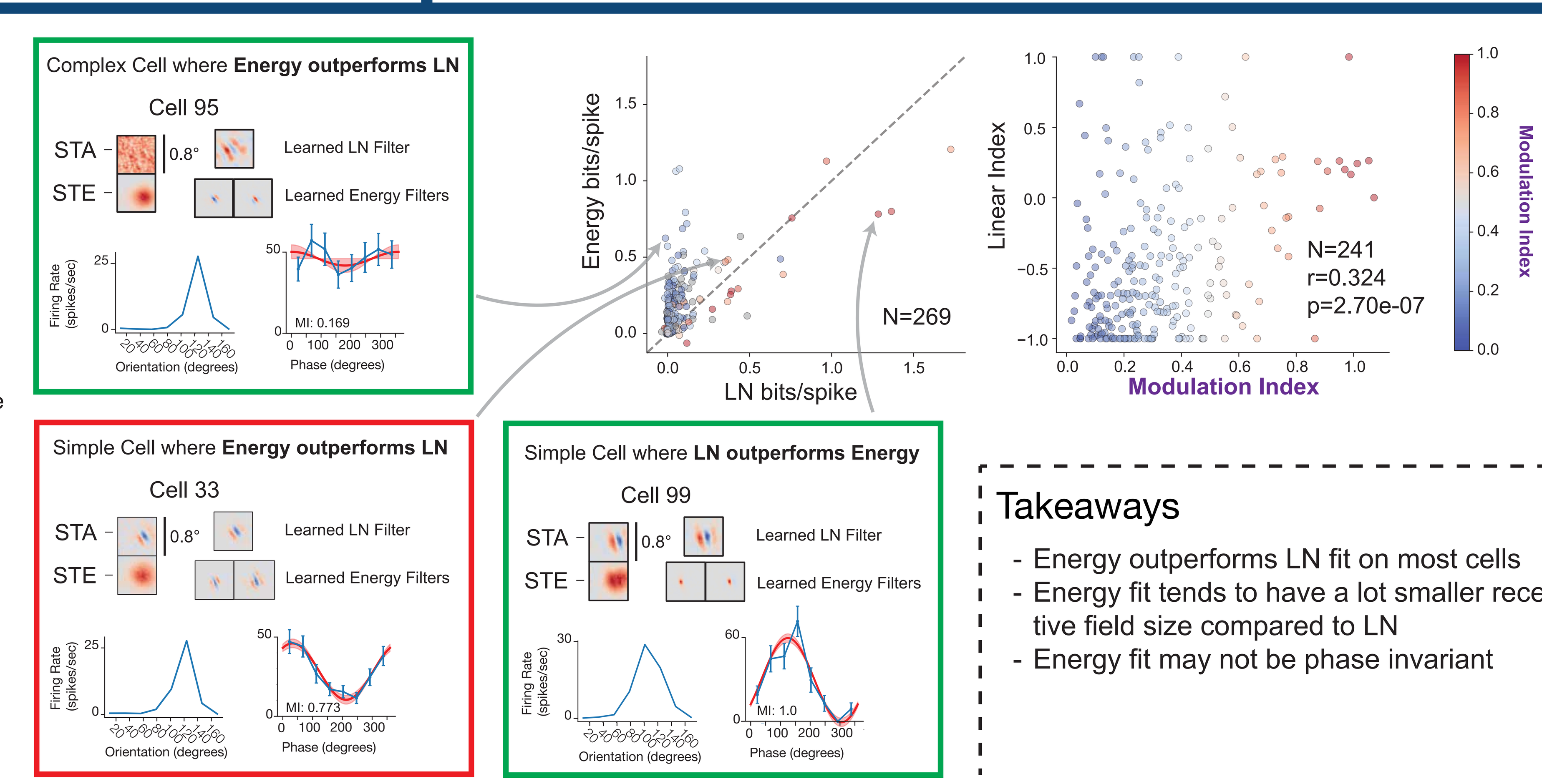
- Most cells (75.6% +- 1.1%) rate modulation can be attributed to eye position trajectory
- Modulation Index, Frequency Tuning, Receptive Field Size predict cell eye movement sensitivity
- Standard models are less invariant to eye movements than real neurons

Fitting standard models



Fitting Details

- Fit to flashed gabor stimulus condition
- LN filter was initialized with Spike Triggered Average (STA)
- Energy Model filters were initialized with first two eigenvectors of Spike Triggered Covariance (STC) matrix
- Kernels are spatiotemporal volumes fit with 18 lags, 4.17 ms bins
- Filters were regularized with smoothness and locality
- Regularization hyperparameters were optimized with RayTune
- Poisson Loss was minimized with LBFGS
- Data was split into Train/Val/Test
- Used hyperparameters at the 80th-percentile validation performance to reduce overfitting.
- Bits/spike is reported on unseen Test set



Takeaways

- Energy outperforms LN fit on most cells
- Energy fit tends to have a lot smaller receptive field size compared to LN
- Energy fit may not be phase invariant

Future directions

- Is energy model learning to be phase sensitive? Fit quadrature pair energy filters
- Condition eye movement sensitivity measures to less than and greater than receptive field size eye movements
- Compare sensitivity of convolutional neural networks to neural code on fixation condition
- Measure eye movement sensitivity of neurons across V1, V2, V4, IT

References

[1] Rucci & Victor, 2015 [2] McFarland et al. 2016
[2] Yates et al., 2023, [3] Ressmeyer et al., 2025

tejasvikothapalli.github.io

yates-lab.github.io

Acknowledgments

This research was supported by the NIH (R00 EY032179 to JLY and R01EY030998 to JFM), a NSF predoctoral fellowship to RAR and a CIVO fellowship to TK.