

Sparse Signal Representation in Digital and Biological Systems

Matthew Guay

University of Maryland, College Park
Department of Mathematics



April 12, 2016

Table of Contents I

1 Preliminaries

2 Compressed Sensing for Electron Tomography

- Background
- Results

3 Sparse Olfactory Coding in the Locust

- Background
- Modeling offset spiking
- KC population sparsity

Introduction

- This work focuses on the implications of signal sparsity in biology, from two very different perspectives.
- **CS-ET**: in what sense are nanometer-scale images of biological structures sparse? Can that sparsity be exploited by compressed sensing?
- **Sparse olfactory coding**: How, and why, do Kenyon cells in locust olfactory processing networks generate sparse sensory codes?

Compressed sensing for electron tomography

- Electron tomography served as a focal point for understanding compressed sensing and sparse mathematical signal processing.
- Signals are vectors in a space of voxel intensities, measurements and representations are linear transforms of the signal.
- **Question:** Can we use compressed sensing to better recover tomograms from undersampled measurement data?

Sparse olfactory coding in locusts

- **Locust olfaction** serves as a common model system for studying **sparse sensory coding** in neuroscience.
- **Lifetime sparsity**: An active Kenyon cell spikes only at onset and possibly offset of a stimulus.
- **Population sparsity**: A small percentage of the Kenyon cell population emit spikes in response to an odor stimulus.
- **Question**: How do locust olfactory population dynamics give rise to this behavior?

Signal representation

- **Signal:** A vector $\mathbf{f} \in \mathbb{R}^M$ for some M .
- A signal model describes the relationships between signals and their measurements and representations.
- **Linear signal model:** Used in the CS-ET project. Signals are represented as linear combinations of basis or frame vectors.
- **Dynamical system signal model:** Used in the locust olfaction project. Signal representations are time-varying states of populations of neurons, whose relationships to the signal are described by nonlinear ODEs.

Sparse signal representation

- A vector \mathbf{x} is **sparse** if its ℓ^0 norm:

$$\|\mathbf{x}\|_0 = \#\{x_i \in \mathbf{x} \mid x_i \neq 0\} \quad (1)$$

is small.

- Sparse linear signal representations aid machine learning by capturing statistical regularities within a class of signals of interest.
- Sparse neural signal representations evidently aid organisms for this and additional reasons.

Compressed sensing

- Compressed sensing (CS): The recovery of a sparse signal $f \in \mathbb{R}^M$ from appropriately-chosen measurements.
- Measurement vectors: A collection of D vectors $\{\varphi_i\}_{i=1}^D \subseteq \mathbb{R}^M$. Each measurement i is $\langle f, \varphi_i \rangle$.
- Representation vectors: A basis or frame $\{\psi_j\}_{j=1}^N$ for \mathbb{R}^M .
- Stack measurement vectors in rows of a measurement matrix $\Phi \in \mathbb{R}^{D \times M}$. Stack representation vectors in rows of a representation matrix $\Psi \in \mathbb{R}^{N \times M}$.

Compressed sensing

- **Sparse signal model:** An *a priori* assumption that $f = \Psi c$ for some $c \in \mathbb{R}^N$ with small ℓ^0 norm.
- In the presence of noise or modeling error, signals are more likely **compressible**: $f \approx \Psi c$ to some suitable level of accuracy.
- **ϵ -compressibility ratio** of a vector x is the proportion of vector components with magnitude greater than $\epsilon \|x\|_\infty$.
- Most existing CS results focus on the cases where Ψ is an orthonormal basis or tight frame, where $f = \Psi^T \Psi f$.

Compressed sensing

- Goal: Given measurements $\mathbf{y} = \Phi \mathbf{f}$, recover \mathbf{f} even if $D \ll M$ as:

$$\mathbf{f}^* = \arg \min_{\mathbf{f} \in \mathbb{R}^M} \|\Psi \mathbf{f}\|_0 \text{ such that } \mathbf{y} = \Phi \mathbf{f}. \quad (2)$$

- Equivalent for some choice of λ to the more useful:

$$\mathbf{f}^* = \arg \min_{\mathbf{f} \in \mathbb{R}^M} \|\Phi \mathbf{f} - \mathbf{y}\|_2^2 + \lambda \|\Psi \mathbf{f}\|_0 \quad (3)$$

- Equation (3) is computationally intractible, we focus on the **convex relaxation** (Basis Pursuit Denoising):

$$\mathbf{f}^* = \arg \min_{\mathbf{f} \in \mathbb{R}^M} \|\Phi \mathbf{f} - \mathbf{y}\|_2^2 + \lambda \|\Psi \mathbf{f}\|_1 \quad (4)$$

Mutual coherence

- Mutual coherence can be used to obtain an upper bound on the number of measurements required for Equation (4) to recover an s -sparse signal f .
- **Mutual coherence:** Given an orthogonal measurement matrix Φ with $\|\varphi_i\|_2 = \sqrt{M}$ for all $i \in [1, D]$, and an orthonormal representation basis Ψ , the mutual coherence of Φ and Ψ is

$$\mu(\Phi, \Psi) = \max_{i,j} |\langle \varphi_i, \psi_j \rangle|. \quad (5)$$

- **Theorem** (Candés, Romberg, 2007): Given D measurements of an s -sparse signal f , Equation (4) recovers f if

$$D \geq C \cdot s \cdot \mu^2(\Phi, \Psi) \cdot \log M, \quad (6)$$

for some small constant C .

The sparse coding hypothesis

- The **sparse coding hypothesis**: Information within a neural population is encoded in a small number of active neurons at each point in time.
- **Population sparsity**: At a fixed point in time, a small proportion of neurons in a population are active.
- **Lifetime sparsity**: A fixed neuron is active for a small proportion of time within an interval of interest.
- Sparse codes minimize overlap between representations of distinct stimuli, useful for associative memory. They are energy efficient, and exploit the statistics of sensory input.
- Found in sensory processing layers across the animal taxonomy, for all sensory modalities.

The sparse coding hypothesis

- Why can sparse codes represent natural stimuli?
- **Hypothesis:** Measurement vectors derived from natural environments lie along a low-dimensional subspace of the ambient measurement space.
- Sparse, overcomplete representations efficiently decompose signal information as a combination of a small number of features.
- In this work, I investigate the mechanisms which drive populations of nonlinear dynamical systems to produce sparse representations of olfactory sense data.

The sparse coding hypothesis

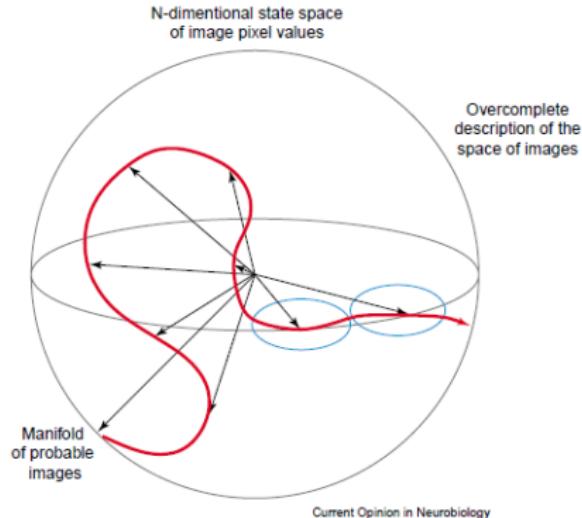


Figure: A stylized depiction of measurement state spaces and the subspace occupied by natural environments. Taken from (Olshausen, Field, 2004).

Table of Contents I

1 Preliminaries

2 Compressed Sensing for Electron Tomography

- Background
- Results

3 Sparse Olfactory Coding in the Locust

- Background
- Modeling offset spiking
- KC population sparsity

Tomography

- **Tomography** - Producing a 3D reconstruction of a specimen by measuring changes in penetrating waves (or particles) which are sent through it.
 - Computed tomography (CT), using X-rays.
 - Electron tomography (ET), using electrons.
- **Electron tomography (ET)** - 3D imaging using electron beams via a transmission electron microscope (TEM) or scanning transmission electron microscope (STEM).

Our ET data

- Each image is a projection of the rotated object, a sequence of images indexed by rotation angle is a **tilt series**.
- **Bright field** (BF) STEM imaging: detectors measure unscattered electrons passing through the specimen.
- **Dark field** (DF) STEM imaging: detectors measure electrons scattered by the specimen.
- Projection contrast comes from density-dependent differences in electron scattering within the specimen.
- Simulated **phantom** datasets were used to compare the efficacy of our CS-ET implementation with similar previous work.

From tomography to Radon transforms

- A beam of n_0 electrons travels along line L through the specimen at each measurement location. Some n electrons pass through undeviated.
- The ratio $\frac{n}{n_0}$ can be related to line integrals of an electron density function $f(\mathbf{x}) : \mathbb{R}^3 \rightarrow \mathbb{R}$ via the Beer-Lambert law:

$$\log\left(\frac{n}{n_0}\right) \propto \int_L f(\mathbf{x}) |d\mathbf{x}| \quad (7)$$

- The function f forms the tomogram recovered from the projection data.

From tomography to Radon transforms

- **Radon transform** - for $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ and any line $L \subseteq \mathbb{R}^2$,

$$Rf(L) = \int_L f(\mathbf{x}) |d\mathbf{x}|. \quad (8)$$

- This space of lines can be parametrized by a normal angle θ and a distance coordinate s :

$$Rf(\theta, s) = \int_{-\infty}^{\infty} f((t \sin \theta + s \cos \theta), (-t \cos \theta + s \sin \theta)) dt.$$

From tomography to Radon transforms

- Parallel beam tomography used in ET decomposes 3D reconstruction into multiple independent 2D reconstruction problems.
- For each plane normal to the rotation axis, tomographic measurements provide samples $\{Rf(\theta_i, s_j)\}_{i \in I, j \in J}$ for some finite sets I, J .
- Measurement limitations make tomogram recovery an ill-posed Radon transform inversion problem.

CS for tomography

- Each sample $Rf(\theta_i, s_j)$ corresponds to a measurement vector $\varphi_{ij} \in \mathbb{R}^D$, all stacked in a measurement matrix Φ .
- Representation operators Ψ used in this work: Identity mapping, discrete DB8 wavelet transform, or the **total variation** operator TV .
- For a 2D discrete image f ,

$$TVf \triangleq \sqrt{\Delta_x^+ f + \Delta_y^+ f}$$

for forward finite $x-$ and $y-$ differences Δ^+ .

Theoretical challenges

- There is little existing theory for CS recovery of signals with sparse images under nonlinear transforms (e.g. TV).
- ET measurement matrices R are deterministic. R and the sparsifying transforms studied here do not possess mutual coherences useful for theoretical analysis of the procedure.

CS-ET mutual coherence

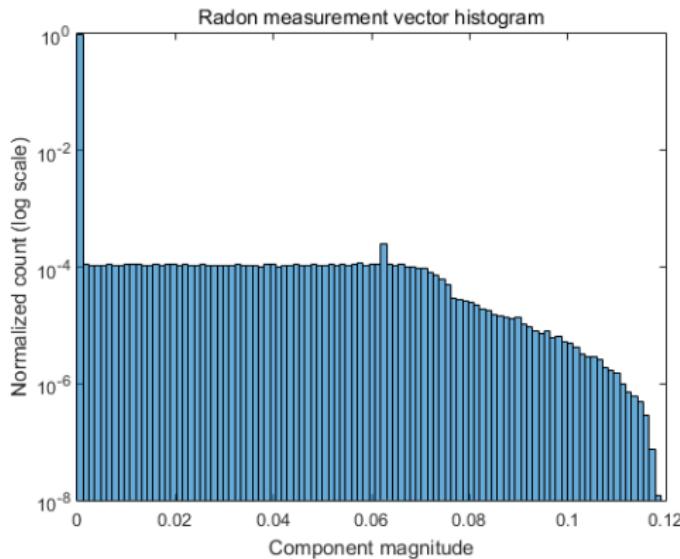


Figure: A histogram of $\langle \varphi_i, \psi_j \rangle$ for $\varphi_i \in R$ and $\psi_j \in I$, an identity matrix. These values are equivalent to the components of Radon transform measurement vectors, taken from a Radon transform of a 256×256 image at angles from -70° to 70° at 5° increments.

CS-ET mutual coherence

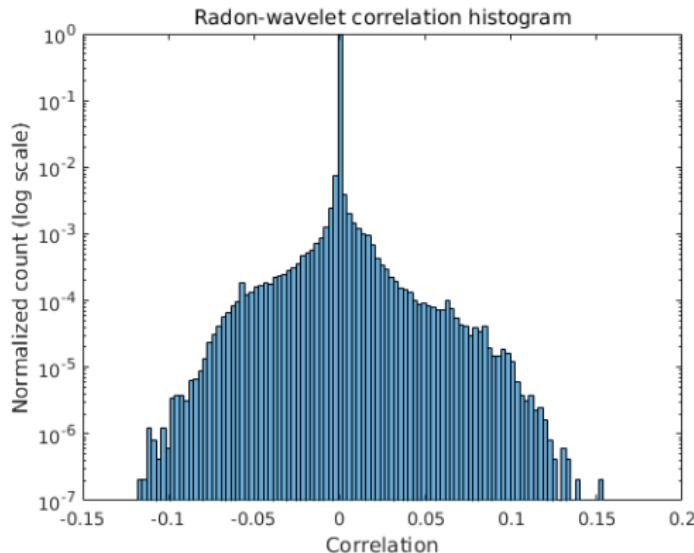


Figure: A histogram of $\langle \varphi_i, \psi_j \rangle$ for $\varphi_i \in R$ and $\psi_j \in W$, the DB8 discrete wavelet transform on \mathbb{R}^{256^2} . Due to computational limitations, only 10% of the possible $\langle \varphi_i, \psi_j \rangle$ combinations were computed.

Numerical methods

- Our CS-ET algorithm computes each $x - z$ slice \mathbf{f}^* of a tomogram as

$$\mathbf{f}^* = \arg \min_{\mathbf{f} \in \mathbb{R}^D} \|R\mathbf{f} - \mathbf{y}\|_2^2 + \lambda_1 \|\mathbf{f}\|_1 + \lambda_2 \|T\mathbf{V}\mathbf{f}\|_1 + \lambda_3 \|W\mathbf{f}\|_1. \quad (9)$$

- R is a digital Radon transform, \mathbf{y} is measurement data, and the λ_i are regularization weights.
- 1024 $x - z$ slices, each approximately 1024×100 .
- We use the **split Bregman** algorithm, a GPU-based library for Radon transform computation, and concurrent computations for multiple $x - z$ slices to solve this problem efficiently.

Tomogram coordinate system

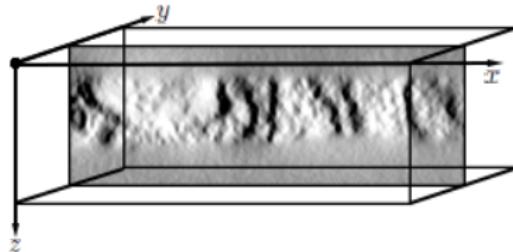
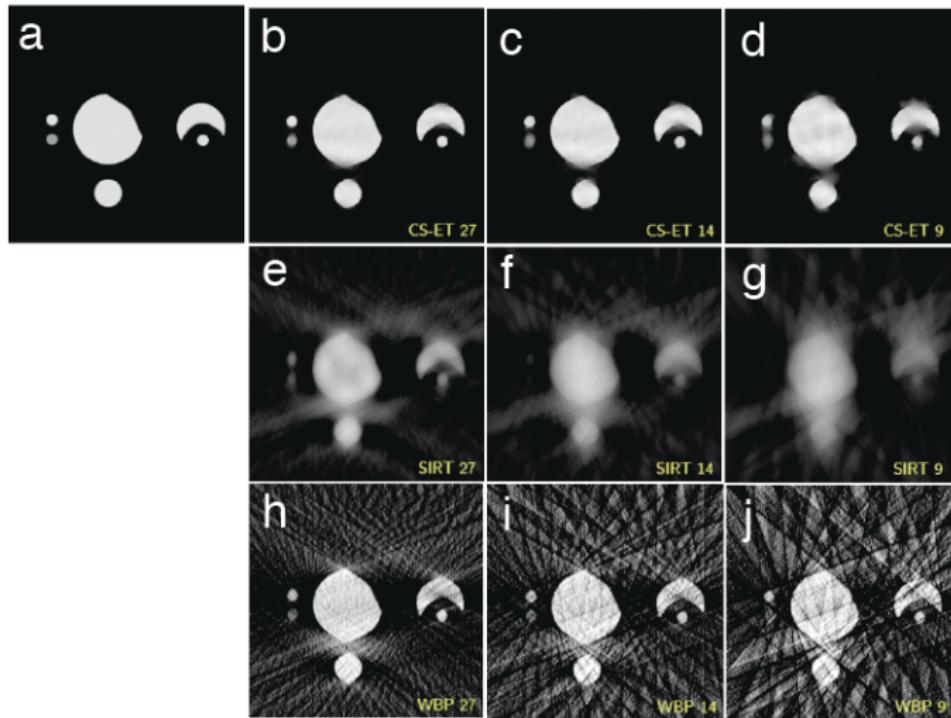
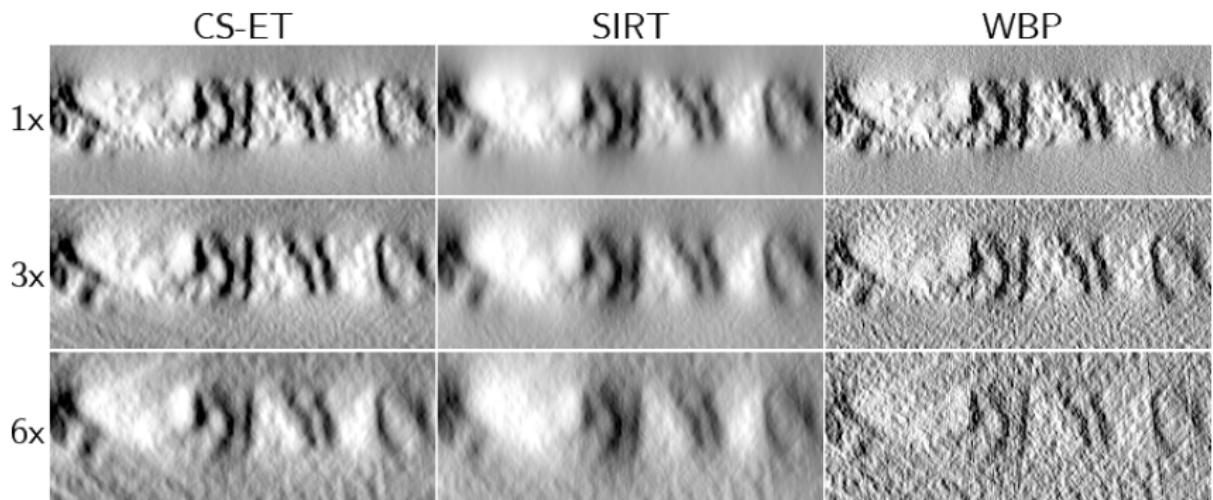
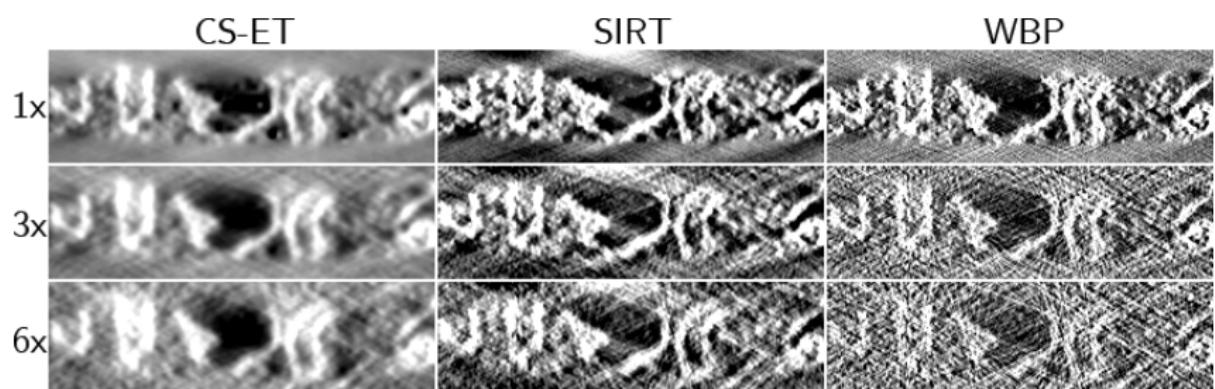


Figure: An illustration of the coordinate system used with 3D tomograms. A tomogram is assembled from independent 2D reconstructions parallel to the $x - z$ plane. An overhead view, parallel to the $x - y$ plane, is useful for visually inspecting tomogram structure.

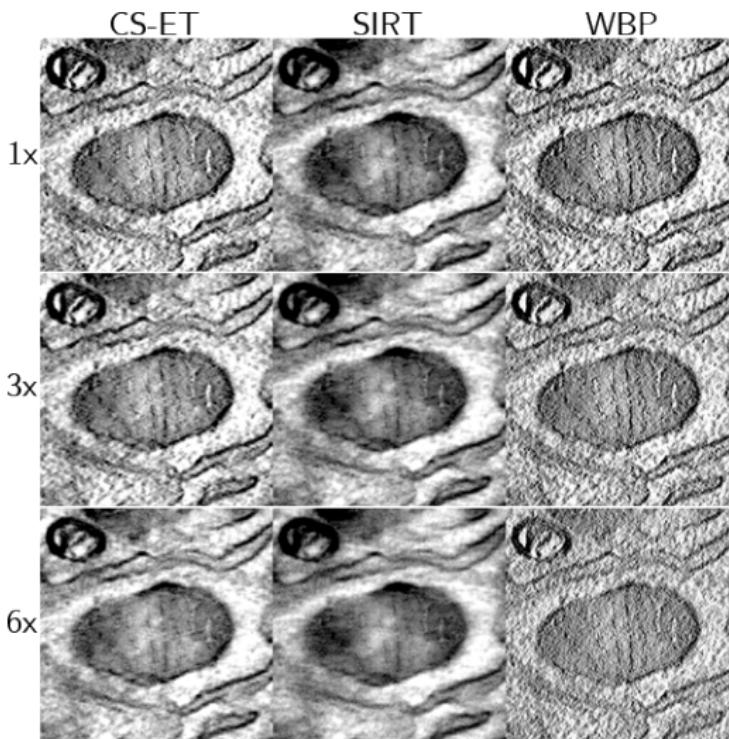
Nanoparticle phantom reconstruction comparison



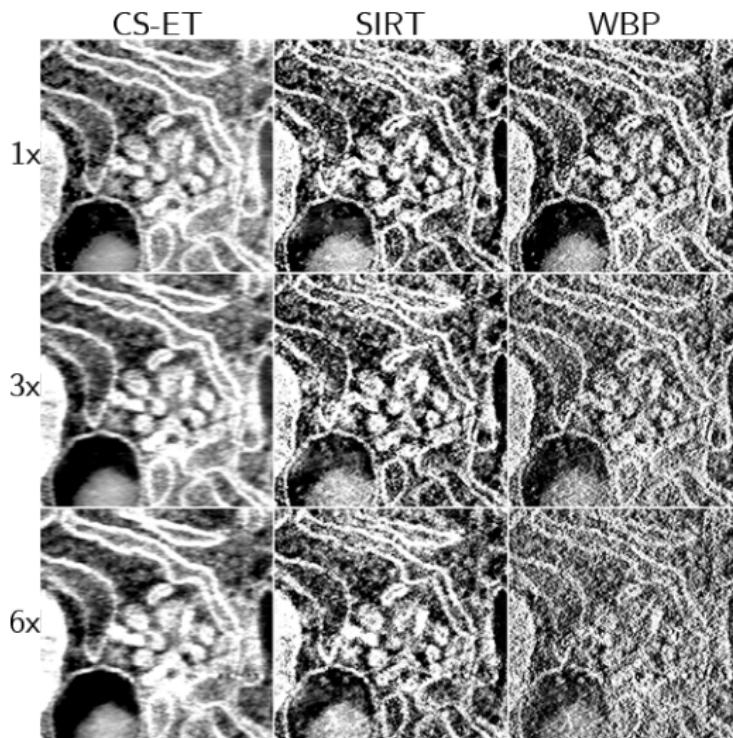
BF STEM reconstruction $x - z$ comparison

DF STEM reconstruction $x - z$ comparison

BF STEM reconstruction $x - y$ comparison



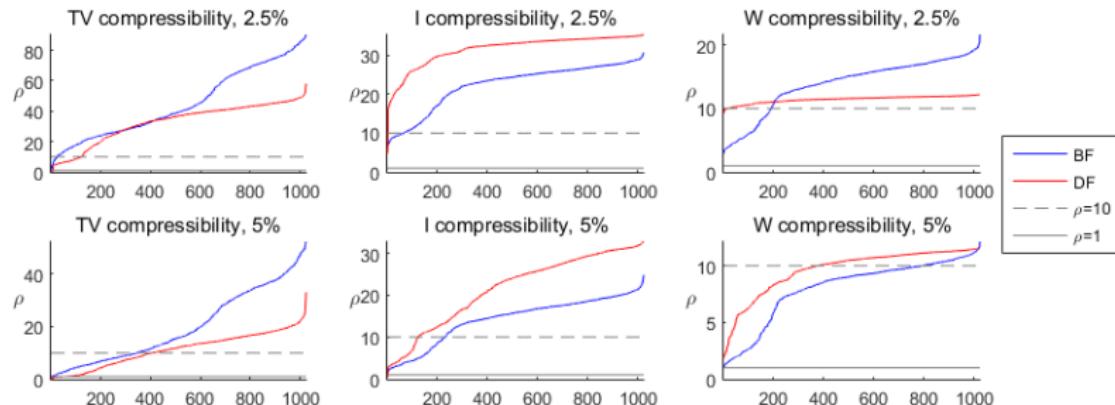
DF STEM reconstruction $x - y$ comparison



CS-ET and structural sparsity

- The $x - z$ slices of the biological STEM datasets are markedly less sparse than the nanoparticle phantom.
- This is a likely source for the disparity in CS-ET performance compared to other reconstruction methods, between the phantom and STEM datasets.
- The correlation between application domain and the structural complexity of specimens is important for determining where CS-ET will be most relevant.

CS-ET and structural sparsity



- BF-STEM and DF-STEM dataset compressibility expressed as ϱ , defined as the STEM datasets' compressibility ratios divided by the nanoparticle phantom's sparsity ratios in each of the three transform domains studied.
- All STEM dataset compressibility values were calculated separately for each $x - z$ slice and ordered by decreasing compressibility.

Table of Contents I

1 Preliminaries

2 Compressed Sensing for Electron Tomography

- Background
- Results

3 Sparse Olfactory Coding in the Locust

- Background
- Modeling offset spiking
- KC population sparsity

Sensory processing

- Long-standing question: How do brains capture, filter, and integrate the information about the environment provided by the senses?
- **Sensory receptors:** neurons whose membrane potentials are directly influenced by external (vision, olfaction) or internal (nociception, proprioception) environmental state.
- Receptors form the first level in a hierarchy of sensory processing centers within brains.

Locust olfaction

- Olfactory systems in insects provide useful models for studying neuron population dynamics.
- Insect olfactory processing systems exhibit complex behavior, but contain relatively few ($\sim 10^6 - 10^7$) neurons and are well-characterized.
- This research focuses on locust olfaction as a model system.
- What network properties cause **Kenyon cells** (KCs) to exhibit both lifetime and population sparsity?

Locust olfaction

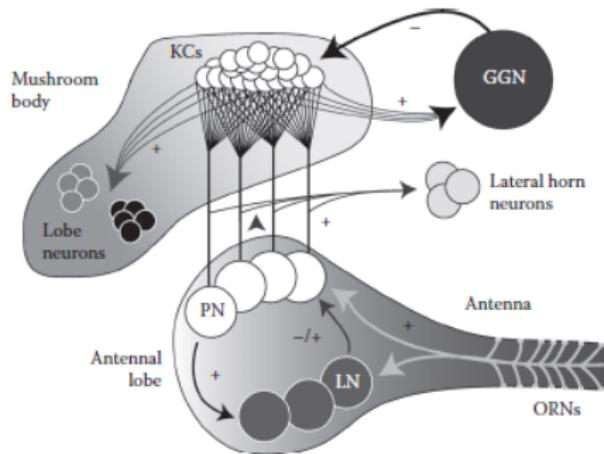


Figure: ~ 50000 olfactory receptor neurons (ORNs) synapse onto ~ 830 projection neurons (PNs) and ~ 300 local neurons (LN). PNs synapse onto ~ 50000 KCs. Used with permission, (DiLorenzo et al., 2014)
Chapter 11.

KC lifetime sparsity

- In the locust, an active KC emits a small number of spikes in response to changes in stimulus identity and concentration, i.e., onset spiking and **offset spiking**.
- **Sensory adaptation** in ORNs drives ORN activity levels towards a baseline during prolonged exposure to a stimulus.
- The time course of this adaptation creates a window of elevated PN activity after a stimulus change.
- Relevant to insect behavior, e.g., for detecting boundaries in odor plumes.

Onset and offset spiking

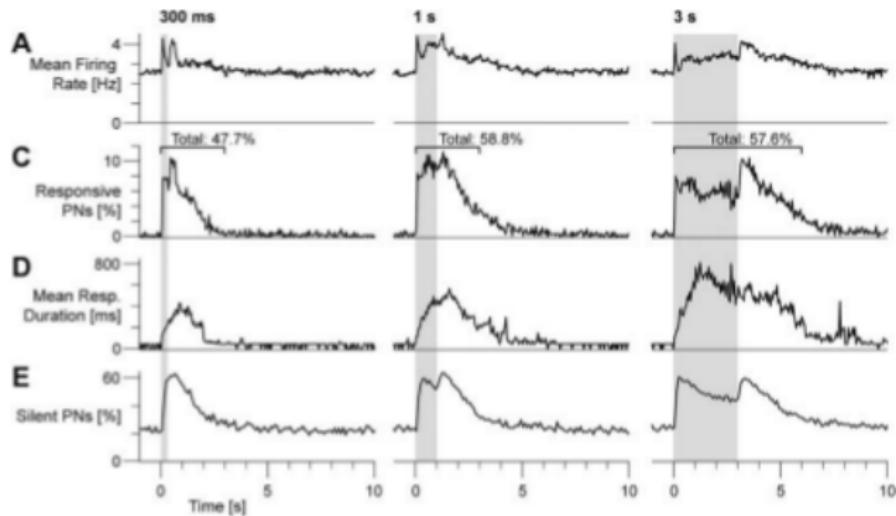


Figure: PN activity data taken from (Mazor, Laurent 2005). All PN population statistics are calculated over time using activity binned into 50ms windows.

Onset and offset spiking

- During stimulation, increased LN→PN feedback inhibition creates oscillatory PN activity which enforces temporal synchrony among active PNs.
- At stimulus onset, increased PN firing rates and synchronous activity patterns lead to onset spiking in KCs.
- LN activity tracks stimulus offset closely, but some PNs continue to spike due to ORNs exhibiting offset activity.
- Stimulus offset creates increased PN firing rates, but decreased temporal coherence.

Modeling offset spiking

- **Question:** What is the relationship between the temporal coherence of a KC's active PN inputs, and the number of active PNs required to elicit a KC spike?
- A suitably-accurate computational model can be used to describe this relationship.
- **Modeling goal:** use a simulation of a KC and its PN synapses to determine how responsive KCs are to different numbers of PN spikes arriving in temporal windows of different lengths.
- How to verify that the model is an accurate reflection of biological KCs?

Model overview

- One KC neuron with a variable number (0-430) of PN→KC synapses.
- N synapses activate randomly within a specified time window $[t_0, t + 0 + \Delta t]$ for varying values of N and Δt .
- KC membrane potential is modeled by a **Hodgkin-Huxley-type** ODE. Synapses are modeled by a standard first-order ODE.
- Implemented in C++ using a 4-step Runge-Kutta (RK4) numerical integrator, time-step $h = 0.03\text{ms}$.

KC model

- KC model: A single-compartment Hodgkin-Huxley-type neuron.
- KC membrane potential V is governed by the an ODE of the form:

$$C \frac{dV}{dt} = -(I_{leak} + I_{int} + I_{syn}) \quad (10)$$

- C is a capacitance constant, I_{leak} is a leakage current, I_{int} is an intrinsic ionic current, I_{syn} is a synaptic current.

KC model

- I_{leak} consists of two components: a “general” leakage current $I_L = g_L(V - E_L)$ and a potassium leakage current $I_{KL} = g_{KL}(V - E_{KL})$.
- Each g is a conductance variable and E a reversal potential.
- I_{int} has five components: $I_{int} = \sum_{i=1}^5 g_i m_i^{M_i} h_i^{N_i} (V - E_i)$, with g_i, E_i as before, $m_i(t)$ and $h_i(t)$ are activation and inactivation variables, and M_i, N_i are experimentally-determined integers.

KC model

- Ionic current conductances:

I_{Na}	$g_{Na} = 26.1 \mu S$	I_K	$g_K = 2.9 \mu S$
$I_{K(Ca)}$	$g_{K(Ca)} = 0.29 \mu S$	I_{Ca}	$g_{Ca} = 0.029 \mu S$
$I_{K,A}$	$g_{K,A} = 0.0145 \mu S$		

- I_{syn} is a sum of individual synaptic currents of the form $g[O](V - E)$, one for each PN → KC synapse. Here, g and E are the same as before, and $[O](t)$ is a proportion of open synaptic channels.

Synaptic model

- $[O](t)$ for each PN→KC synapse is updated as

$$\frac{d[O]}{dt} = \alpha(1 - [O])[T] - \beta[O]. \quad (11)$$

- $[T](t)$ measures transmitter concentration, α is a synaptic current rise rate parameter, and β is a synaptic current decay rate parameter.
- Synapse strength g is not uniform across PN→KC connections.

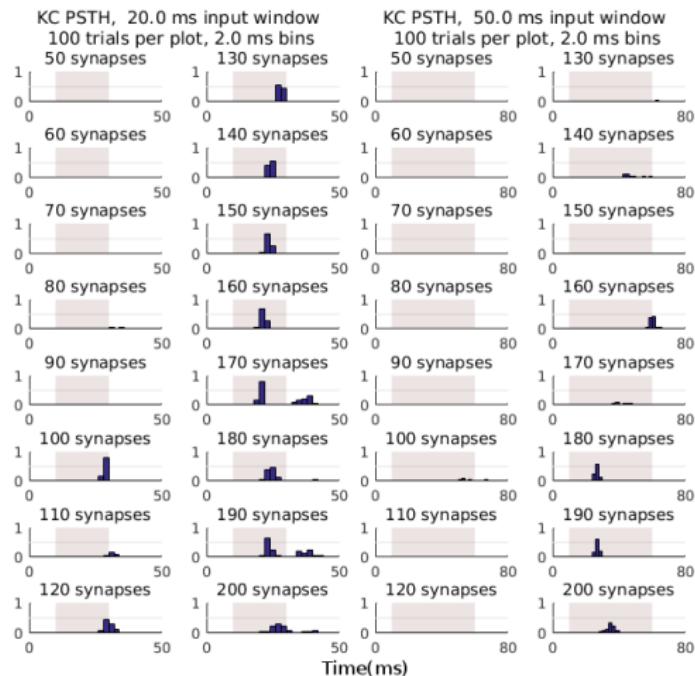
Synaptic model

- To test the effect of coordinated PN spiking on the KC, each of N synapses is set to activate at a time drawn uniformly at random from the interval $[t_0, t_0 + \Delta t]$.
- Biological PNs exhibit nonzero baseline firing rates ($\sim 2.5\text{Hz}$) which may be important for tuning the KC's responsiveness to coordinated spikes.
- Model this by adding $415 - N$ synapses with no coordinated spiking time, to all 415 synapses assign random spiking events with exponentially-distributed interarrival times ($\lambda = 0.0025$).

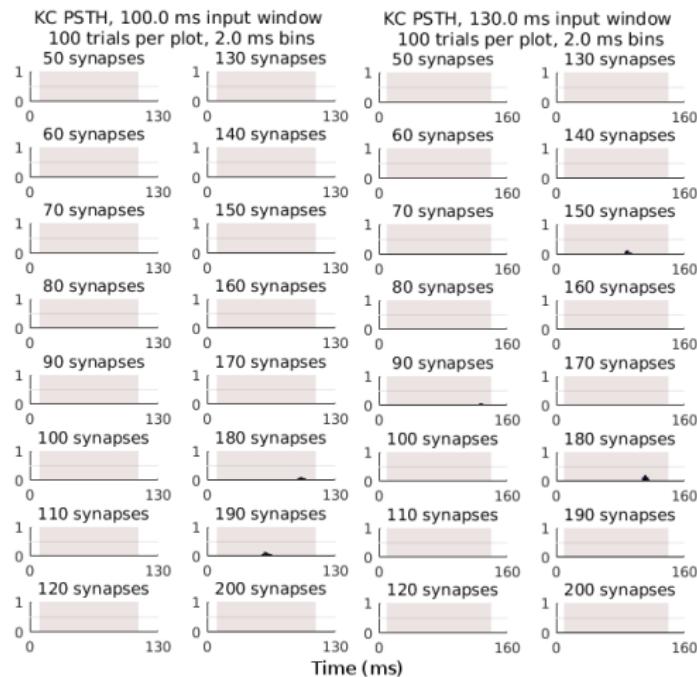
Simulation protocol

- For each N in $\{50 : 10 : 200\}$ and each Δt in $\{10 : 10 : 60\} \cup \{100 : 15 : 400\}$, simulate N synapses with activation times in $[t_0, t_0 + \Delta t]$. Run the simulation for $K = 100$ trials.
- Using the K trials, for each parameter configuration construct a KC **peristimulus time histogram** (PSTH).
- PSTH: For each trial, bin spike counts in time, then average the binned counts across all trials.

KC PSTHs



KC PSTHs



Model validation

- Computational modeling indicates that the 150-200 PNs spiking across a 150-300ms interval (like at stimulus offset) are unlikely to trigger a KC spike.
- Is this modeling error? How do we assess the biological relevance of this model?
- **Goal:** Model should conform with known activity statistics for PN→KC interactions and KC behavior.
- Focuses so far: parameter selection, synaptic conductance distribution, KC resting potential and firing threshold, KC membrane time constant.

Explaining KC population sparsity

- Useful KC population analysis is difficult to analyze as a large nonlinear ODE model.
- To what extent can KC population activity be explained more simply?
- **Goal:** Produce activity statistic distributions for KCs, in a simplified PN and KC network.
- Choose statistics to explain how KC activity is sparse, and how KC activity is a representation of sensory information.

Binary time series model

- **Simple model:** KC activity is binned over time. Within each time bin, use a binary active/silent model for each PN and KC.
- A KC is connected to $K = 415$ of the 830 PNs. The PN target set is chosen uniformly at random from the possible subsets of PNs.
- A KC is active in a time bin if $\tau = 100$ or more PNs are active.

KC population sparsity

- Derive that

$$\varrho(M) = P(\text{active KC} | M \text{ active PNs}) = \sum_{k=\tau}^M \frac{\binom{M}{k} \binom{830-M}{K-k}}{\binom{830}{k}}. \quad (12)$$

- For this model, 150 active PNs marks a transition from a low probability of KC activity to a high probability.
- For a fixed M and 50000 KCs, we compute that

$$P(s \text{ total KC spikes}) = \binom{50000}{s} \varrho(M)^s (1 - \varrho(M))^{50000-s}. \quad (13)$$

KC population sparsity

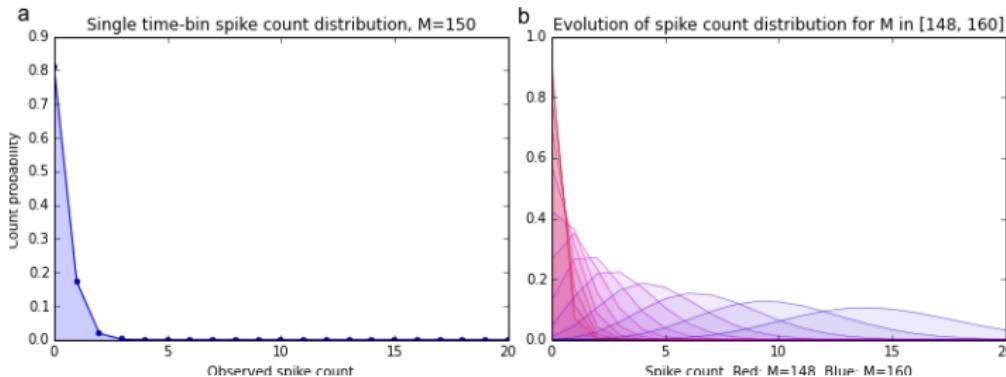


Figure: Single time bin plot of $P(s$ total KC spikes) for several values of M , the number of active PNs.

KC response distinguishability

- Define B_i as the binary time series of a KC population in response PN population time series A_i .
- **Goal:** Compute the distribution of $\|B_1 - B_2\|_1$, for two PN activity series A_1, A_2 .
- A_1 and A_2 have a fixed number of active PNs in each time bin, chosen uniformly at random from the population.
- **Result:** $\|B_1 - B_2\|_1$ is large with high probability, but specifics are dependent on a number of parameters.

End

Thank you!

Works cited |



Emmanuel Candes and Justin Romberg.

Sparsity and incoherence in compressive sampling.

Inverse problems, 23(3):969, 2007.



Patricia M DiLorenzo and Jonathan D Victor.

Spike timing: mechanisms and function.

CRC Press, 2014.

Works cited II



Ron A Jortner, S Sarah Farivar, and Gilles Laurent.

A simple connectivity scheme for sparse coding in an olfactory system.

The Journal of neuroscience, 27(7):1659–1669, 2007.



Ofer Mazor and Gilles Laurent.

Transient dynamics versus fixed points in odor representations by locust antennal lobe projection neurons.

Neuron, 48(4):661–673, 2005.

Works cited III

-  Bruno A Olshausen and David J Field.
Sparse coding of sensory inputs.
Current opinion in neurobiology, 14(4):481–487, 2004.
-  Javier Perez-Orive, Maxim Bazhenov, and Gilles Laurent.
Intrinsic and circuit properties favor coincidence detection for decoding oscillatory input.
The Journal of neuroscience, 24(26):6037–6047, 2004.

Appendix: Model parameter selection

- (Perez-Orive et al. 2004) describes a Hodgkin-Huxley model of the locust KC. Our model is based on this one.
- Errors in the paper required communication with Maxim Bazhenov to obtain their model source code.
- Our model uses an I_{Ca} current and calcium dynamics described in the source code, differing from the paper.
- Leakage conductance g_L has been increased from 2.9×10^{-3} to 2.9×10^{-2} , fixing the KC resting potential and making the KC less quiescent.

Appendix: Synaptic conductance distribution

- PN→KC synaptic conductances are not uniform across the population. Their distribution may be estimated from the EPSP amplitude distributions recorded in (Jortner et al. 2007).
- I created a conductance distribution function matching this distribution, then computed a simulated EPSP amplitude distribution to verify its match with the Jortner et al. data.

Appendix: Synaptic conductance distribution

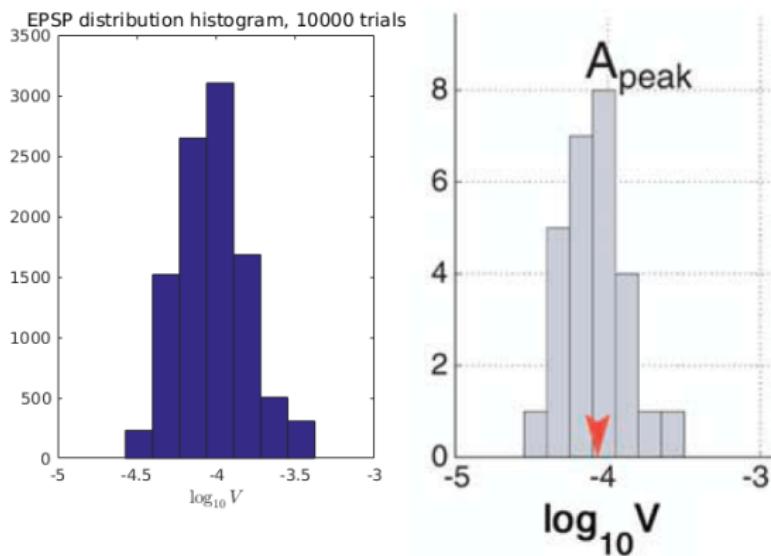


Figure: A comparison of a simulated peak EPSP distribution (left) and the equivalent data from (Jortner et al., 2007) (right).

Appendix: KC firing threshold

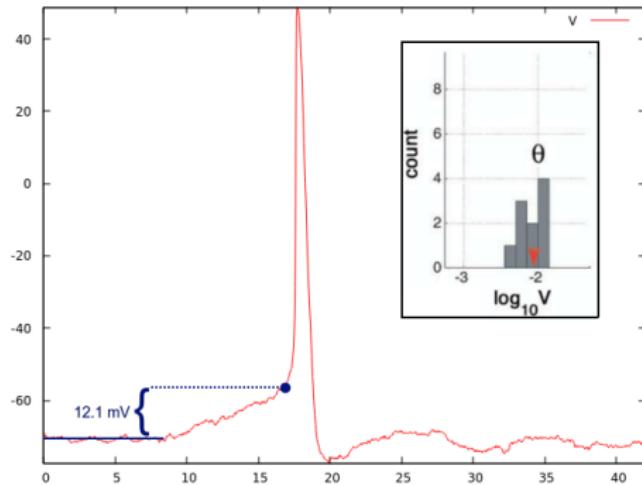


Figure: A comparison of a simulated KC's firing threshold with a firing threshold distribution measured in (Jortner et al., 2007). A neuron's firing threshold is defined here as the difference between resting potential and the point with largest second derivative during a spike response.

Appendix: KC membrane time constant

- **Membrane time constant:** Amount of time required for a neuron to transition $(1 - 1/e) \approx 63.2\%$ of the distance from a membrane potential depolarized by a square current pulse, back to equilibrium.

Current (nA)	0.025	0.05	0.1	0.2	0.25
STC (ms)	6.89	6.83	6.77	6.89	7.37

Table: Simulated KC membrane time constant measurements from square current pulses of several amplitudes. Comparisons with biological experiments are forthcoming.

Appendix: Future validation work

- Current validation procedures give little explicit comparison between simulated ionic current behaviors and their biological counterparts.
- A **voltage clamp** experiment gives more detailed information about the magnitude of ionic current flowing through a neuron at a variety of membrane potentials.
- Voltage clamps can be simulated - code is currently written, awaiting electrophysiological data to use for comparison.