**Thesis outline:**

**Abstract:** Need of the thesis, goal, hypothesis, results, Interpretation (future things??)

**Introduction:**

Introduction to problem,

Some background,

Recent works, comparison of works,

Knowledge gap,

Methods which is similar but why can’t be applied here

**Data:**

Source, what it contains, see it in [C:\Users\Smit3\Downloads\Mesoscale-Activity-Analysis2\Report\version3.pdf](file:///C:\Users\Smit3\Downloads\Mesoscale-Activity-Analysis2\Report\version3.pdf)

Maybe try to get more metadata about

**Data Engineering:**

Data curation part, all the operation and all functional operations done on the data

**Data cleaning:**

see it in [C:\Users\Smit3\Downloads\Mesoscale-Activity-Analysis2\Report\version3.pdf](file:///C:\Users\Smit3\Downloads\Mesoscale-Activity-Analysis2\Report\version3.pdf)

maybe include queries, like good bad trials, CCF overlaps.

**Data Processing:**

**Methodology:**

**Firing Rate:**

**Spike statistics:**

**Dimentionality reduction:**

**Coding direction**

**PCA**

**Multimodal analysis:**

**CCG:**

**Autocorrelograms:**

**Tau,intrinsic timescale:** Diminishing Value,

**Jitter:**

**Fluorescence Analysis:**

**CCF-CCG:** Hypothesis,

**p-value:** Statistical Methods, CCG peak, Coincidence, there are other kind of simulation test in [C:\Users\Smit3\Downloads\Mesoscale-Activity-Analysis2\Literature\Thesis\amarasingham-et-al-2012-conditional-modeling-and-the-jitter-method-of-spike-resampling.pdf](file:///C:\Users\Smit3\Downloads\Mesoscale-Activity-Analysis2\Literature\Thesis\amarasingham-et-al-2012-conditional-modeling-and-the-jitter-method-of-spike-resampling.pdf)

**Content**

**Abstract**

from [C:\Users\Smit3\Downloads\Mesoscale-Activity-Analysis2\Report\version3.pdf](file:///C:\Users\Smit3\Downloads\Mesoscale-Activity-Analysis2\Report\version3.pdf)

and new conclusion and methods according to recently done work.

**Introduction**

from [C:\Users\Smit3\Downloads\Mesoscale-Activity-Analysis2\Report\version3.pdf](file:///C:\Users\Smit3\Downloads\Mesoscale-Activity-Analysis2\Report\version3.pdf)

and rest of the introduction and Related work are In the Conversation of “SMIT THESIS”

Need to Do more Literature review >15papers

Summary of Papers

**Neurons as Canonical Correlation Analyzers (Pehlevan et al.)**

Summary: Pehlevan et al. propose that neural networks can function as Canonical Correlation Analyzers (CCA), aligning activity patterns in response to correlated inputs. CCA identifies shared variance between two sets of variables, revealing correlated patterns across neuron groups.

Relevance: Given your focus on inter-regional communication, CCA provides a promising tool for capturing shared activity patterns between ALM and Thalamus neurons. This aligns with your project’s aim of understanding correlated activity, as CCA can identify functional connectivity based on shared responses.

Potential Use: CCA is well-suited to your project for analyzing neural data from multiple regions to detect patterns of inter-regional interaction. It complements PCA by not just reducing dimensionality but specifically correlating activity across regions, making it highly applicable in examining ALM-Thalamus connectivity.

**The Brain and Its Time: Intrinsic Neural Timescales are Key for Input Processing and** **A Hierarchy of Intrinsic Timescales across Primate Cortex (Murray et al.) is complementing each other Details are in the Chat.**

**Dimensionality Reduction for Large-Scale Neural Recordings (Cunningham & Yu, 2014)**

**Potential Use**: This foundational work validates the application of PCA in your project. Additionally, it suggests that exploring other reduction techniques (e.g., t-SNE or manifold learning) could add depth to your analysis, especially if PCA does not fully capture the structure in your dataset.

**Steinmetz, N. A., Zatka-Haas, P., Carandini, M., & Harris, K. D. (2019). "Distributed coding of choice, action, and engagement across the mouse brain." Nature, 576(7786), 266-273.**

Summary: This paper explores distributed neural activity across multiple brain regions involved in choice and action in mice. It provides insights into neural coding across brain regions and demonstrates how different areas contribute to behavior-related processing.

Relevance: This study could support your focus on cross-regional neural dynamics, especially for understanding how ALM and Thalamus interact in movement planning and decision-making.

Potential Use: You could use Steinmetz et al.’s findings to contextualize your focus on inter-regional interactions in movement planning tasks, particularly in relation to distributed coding of decision-related activity.

**Stringer, C., Pachitariu, M., Steinmetz, N., Reddy, C. B., Carandini, M., & Harris, K. D. (2019). "Spontaneous behaviors drive multidimensional, brainwide activity." Science, 364(6437), eaav7893.**

Summary: Stringer et al. examine spontaneous neural activity across the brain, showing that behaviors can activate complex, multi-dimensional patterns in neural responses.

Relevance: The study highlights the relevance of spontaneous and behavior-driven activity patterns in the brain, underscoring the importance of multidimensional analysis techniques like PCA for uncovering complex neural dynamics.

Potential Use: This paper supports your use of PCA and Coding Direction for capturing complex, behaviorally relevant neural patterns and provides context for interpreting multidimensional activity patterns.

**Kobak, D., & Berens, P. (2019). "The art of using t-SNE for single-cell transcriptomics." Nature Communications, 10(1), 1-14.**

Summary: While focused on single-cell transcriptomics, this paper provides a comprehensive overview of t-SNE (t-Distributed Stochastic Neighbor Embedding), a powerful dimensionality reduction technique that could complement PCA.

Relevance: Kobak and Berens provide insights into effectively using t-SNE for high-dimensional data, which might be relevant if you want to compare or complement PCA with another method.

Potential Use: t-SNE could be applied to your dataset to create low-dimensional representations of neural activity patterns. Adding t-SNE as an alternative to PCA may provide a different perspective on clustering behavior-related neural patterns.

**Cunningham, J. P., & Byron, M. Y. (2014). "Dimensionality reduction for large-scale neural recordings." Nature Neuroscience, 17(11), 1500-1509.**

Summary: This foundational review outlines various dimensionality reduction methods applicable to neural recordings, providing comparisons and best practices for neural data interpretation.

Relevance: This paper aligns directly with your use of PCA and Coding Direction, offering theoretical support for the choice of dimensionality reduction techniques in high-dimensional neural data.

Potential Use: This can strengthen the justification for using dimensionality reduction in your analysis, particularly in your exploration of how neurons across brain regions encode movement planning tasks.

**Fries, P. (2005). "A mechanism for cognitive dynamics: neuronal communication through neuronal coherence." Trends in Cognitive Sciences, 9(10), 474-480.**

Summary: Fries proposes that neural coherence (synchronization of oscillatory patterns) enables communication across brain regions, providing a framework for understanding functional connectivity dynamics.

Relevance: While your primary focus is on Cross-Correlogram analysis, coherence-based connectivity analysis could offer a complementary view, particularly in understanding how ALM and Thalamus may coordinate their activity in a time-locked manner.

Potential Use: You may cite Fries’ theory to discuss potential underlying mechanisms of inter-regional connectivity, even if you’re not using coherence analysis, as it provides a theoretical basis for understanding synchronous neural dynamics.

**Hasson, U., Chen, J., & Honey, C. J. (2015). "Hierarchical process memory: Memory as an integral component of information processing." Trends in Cognitive Sciences, 19(5), 304-313.**

Summary: This paper discusses the role of intrinsic timescales in hierarchical memory processes across the brain. Hasson et al. propose that regions with longer timescales play crucial roles in sustaining information over time, relevant to tasks involving memory.

Relevance: This theoretical framework could support your study’s analysis of temporal integration in ALM and Thalamus during memory-guided tasks.

Potential Use: You can use this paper to justify the relevance of intrinsic timescales in analyzing memory-related tasks and decision-making, particularly as you investigate how different timescales support the maintenance of information in ALM.

**Chaudhuri, R., Knoblauch, K., Gariel, M. A., Kennedy, H., & Wang, X. J. (2015). "A large-scale circuit mechanism for hierarchical dynamical processing in the primate cortex." Neuron, 88(2), 419-431.**

Summary: Chaudhuri et al. investigate hierarchical dynamics in the primate cortex, showing that different areas contribute uniquely to a network's temporal properties. They emphasize the role of recurrent circuits in maintaining long-term activity.

Relevance: This work aligns closely with your study’s exploration of temporal dynamics in ALM and Thalamus, suggesting that such hierarchical temporal properties could play a role in your recorded regions.

Potential Use: This paper could serve as a theoretical basis for examining temporal dynamics in your study, helping to contextualize the role of ALM and Thalamus in maintaining task-relevant information over time.

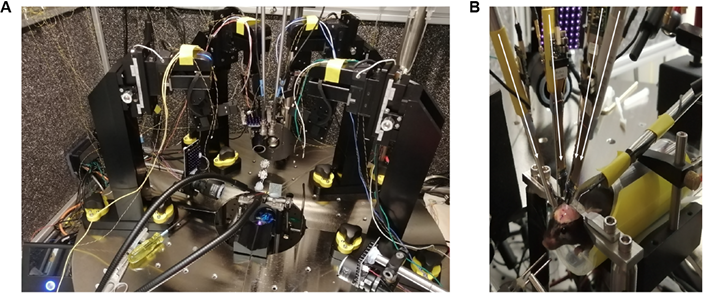
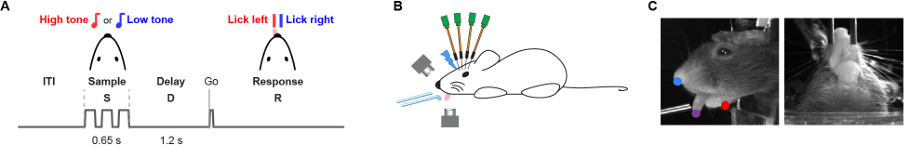
**Related work/Background (knowledge gap, Contribution)**

**Data**

* **Data acquisition/Source/availability/Dataset/Experiment Settings**

Starting and bit from [C:\Users\Smit3\Downloads\Mesoscale-Activity-Analysis2\Report\version3.pdf](file:///C:\Users\Smit3\Downloads\Mesoscale-Activity-Analysis2\Report\version3.pdf)

And other following bits are from Chen et al.[C:\Users\Smit3\Downloads\Mesoscale-Activity-Analysis2\Literature\Chenmultiregional.pdf](file:///C:\Users\Smit3\Downloads\Mesoscale-Activity-Analysis2\Literature\Chenmultiregional.pdf)



**Animals and surger**y

This study is based on data from 28 mice (Table 1), including twenty five VGAT-ChR2 EYFP (Jackson laboratory, JAX #014548), one C57BL/6J (JAX #000664), one Sst IRES-Cre (JAX #013044) crossed with reporter mouse Ai32 (Rosa26-LSL-ChR2-EYFP, JAX #012569), and one Emx1-Cre (JAX #005628) crossed with R26-LNL-GtACR1 Fred-Kv2.1 reporter mouse (JAX #033089). See Table 2 for recordings made in each mouse. All procedures were in accordance with protocols approved by the Janelia Research Campus Institutional Animal Care and Use Committee. Detailed information on water restriction has been published (Guo et al. 2014). Mice were housed in a 12:12 reverse light:dark cycle and performed behavioral tasks during the dark phase. Behavioral sessions lasted 1 to 2 h where mice received all their water (range, 0.3 to 1.5 mL). All surgical procedures were carried out aseptically. Mice were implanted with a titanium headpost and single housed. Recording well and craniotomy preparation for electrophysiology in head-restrained awake mice have been described in detail (dx.doi.org/10.17504/protocols.io.9a8h2hw). Buprenorphine (0.1 mg/kg, IP injection) was used for postoperative analgesia and Ketoprofen (5mg/kg, subcutaneous injection) 40 The copyright holder for this preprint bioRxiv preprint doi: https://doi.org/10.1101/2023.03.01.530520 ; this version posted March 2, 2023. (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license . was used at the time of surgery and postoperatively for two days. After head post implantation, all mice were allowed to recover for at least 3 days with free access to water before the start of water restriction. Craniotomies for recording were made after behavioral training.

**Behavior and video tracking**

Mice performed an auditory delayed response task (Figure 1A) (Inagaki et al. 2018). The instruction stimuli during the sample epoch were pure tones played at one of two frequencies (3 kHz or 12 kHz). Each tone was played three times for 150 ms with 100 ms inter-tone intervals. The sample epoch was followed by a 1.2 s delay epoch. An auditory ‘Go’ cue (carrier frequency of 6 kHz with 360 Hz modulating frequency, 0.1 s duration) indicated the end of the delay epoch. Licking early during the sample/delay epoch triggered a replay of the epoch. During the response epoch (answer period: 1.5 s) mice reported the instruction by licking one of the two lick ports. Licking (consumption period: 1.5 s) the correct lick port triggered a small water reward (0.1 ~ 0.2 ul). Licking the incorrect lick port triggered a timeout (1-3 s). After mice stopped licking for 1.5 s, the trial ended, followed by a 250 ms inter-trial-interval. Early lick trials and no response trials were excluded for analysis. Two CMOS cameras (CM3-U3-13Y3M, FLIR) were used to track orofacial movements of the mouse under IR light illumination (940 nm LED). The cameras were equipped with 4-12 mm focal length lenses (12VM412ASIR, Tamron) and a pixel resolution of 71 µm. High-speed videos from a side view and a bottom view (Figure 1C) were acquired at 300 Hz using software FlyCapture (TELEDYNE FLIR) and trained DeepLabCut (Mathis et al. 2018) to track the movement of tongue, jaw and nose (Figure 1I).

**Multi-regional electrophysiological recordings**

We designed a flexible manipulator system for multiple Neuropixels probe insertions (https://www.janelia.org/open-science/manipulator-system-for-multiple-neuropixels probe-recordings). Three to four small craniotomies (diameter, 1 ~ 1.5 mm) were made over target brain areas one day before the first recording session. Details of recordings made with multiple Neuropixels probes in head-restrained behaving animals have been published (dx.doi.org/10.17504/protocols.io.8tphwmn). Target regions, insertions angles, probe types, etc are summarized in Table 2. All coordinates are given with respect to Bregma (AP, anterior-posterior; ML, medial-lateral; DV, dorsal-ventral). Extracellular spikes were recorded using Neuropixels 1.0 probes, 2.0 single-shank probes, and 2.0 multi-shank probes (J. J. Jun 2017; Steinmetz et al. 2021) (Table 2). Three to seven recordings were made from each craniotomy. Recording depth was inferred from manipulator readings (Sensapax uMp manipulators). After completion of probe insertion, brain tissue was allowed to settle for several minutes before recording. During recordings the craniotomies were immersed in saline. Probes were configured and data were visualized and streamed to disk using the open source software package SpikeGLX (https://billkarsh.github.io/SpikeGLX/). Data from up to five simultaneously recorded probes were acquired at 30 kHz. Sync waves (0.5 s duty cycle TTL pulses) were recorded across probes and on auxiliary channels for synchronization.

* **Data Processing/Engineering/Curating/Cleaning**

**From** [C:\Users\Smit3\Downloads\Mesoscale-Activity-Analysis2\Report\version3.pdf](file:///C:\Users\Smit3\Downloads\Mesoscale-Activity-Analysis2\Report\version3.pdf)

But add the actual names of techniques like

For Data org/acq : File handling : pynwb and h5py, DANDI API CLI(Mostly in above subsection),

Data Transformation: Defining/segmenting in epochs, Spike binning (kernel smoothing)

**Consistent Referencing**: Ensure consistent referencing for each neural unit, electrode, and anatomical region. Keeping track of these references across sessions is important for multi-region analyses. This can be facilitated by assigning unique identifiers to electrodes and neurons and using metadata files or lookup tables.

**Anatomical Filtering**: Focus on data from specific brain regions, such as ALM and Thalamus, by filtering units or channels based on their anatomical metadata. This is typically specified in tables or metadata files in the NWB dataset

**Temporal Alignment:** Align the start and end times of each epoch (e.g., stimulus, delay, and go) to ensure temporal consistency. Any session where epochs are not sequential or contain time gaps should be flagged for potential removal or correction.

**Trial Exclusion Criteria**: Trials with abnormal durations (outliers) or where epoch timings are inconsistent (e.g., delay periods lasting outside of the expected 1–1.3 seconds) should be excluded. Establishing a maximum trial duration (e.g., 5.3 seconds in our case) helps maintain uniformity in trial analysis.

**Epoch Outliers**: distribution of duration(check very high/low duration nd remove them)

**Spike Outliers:** Spike with low firing rate

Additional Queries as per Requirements and further analysis??

GPT SUGGESTED INFO: (REQUIRED?????)

**Metadata Documentation**

* **Detailed Metadata**: Document all preprocessing steps and session-level details (e.g., electrode coordinates, anatomical labels, session timestamps). This metadata is essential for reproducibility and for multi-region analyses that require consistency across sessions and subjects.

**Automated Checks**:

Implement automated checks to verify key data characteristics after each processing step, such as trial durations, firing rate distributions, and epoch alignment. These checks can flag abnormalities early, ensuring that only high-quality data progresses through the analysis pipeline.

 **Manual Review**: Visualize key data characteristics (e.g., spike rasters, firing rate histograms, epoch distributions) to manually spot-check for any anomalies that automated checks might miss. Reviewing a subset of trials per subject can also help confirm that processing steps are being consistently applied.(THIS IS SOMETHING I DID)

By rigorously applying these methods, you’ll be able to maintain high data quality, ensuring that your subsequent analyses are both reliable and interpretable.

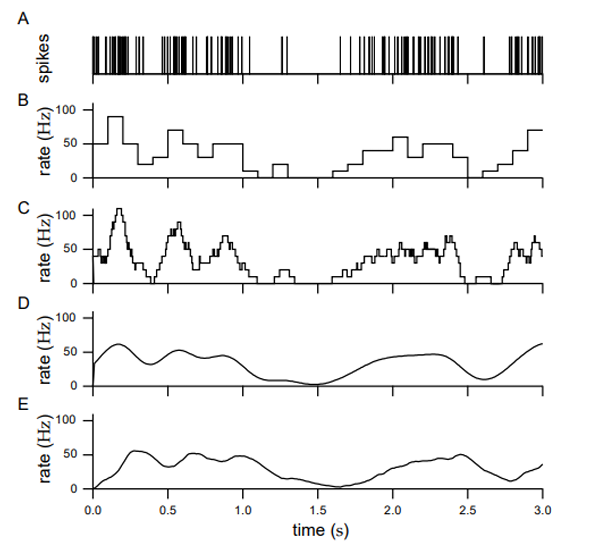
**Methodology/Data Analysis:**

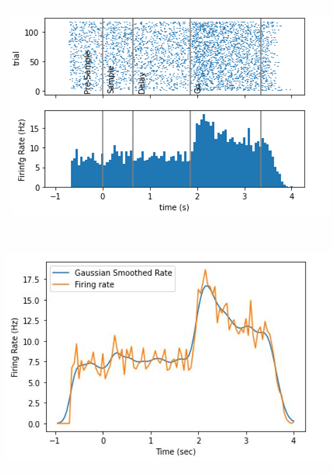
* **Preprocessing:**

**Firing Rate and Kernel smoothing**:

Spike Statistics:

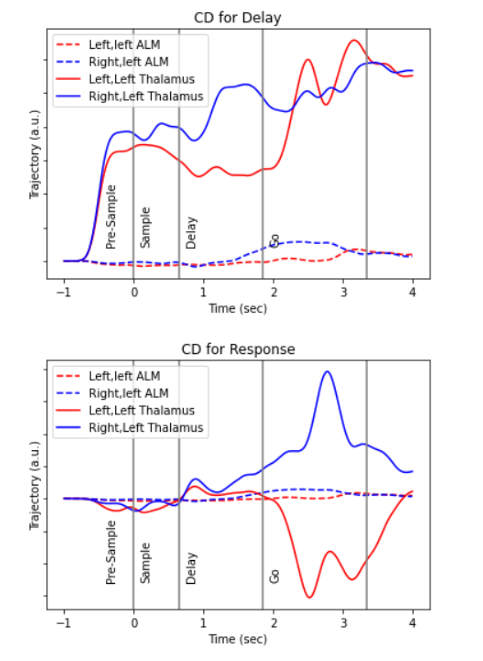
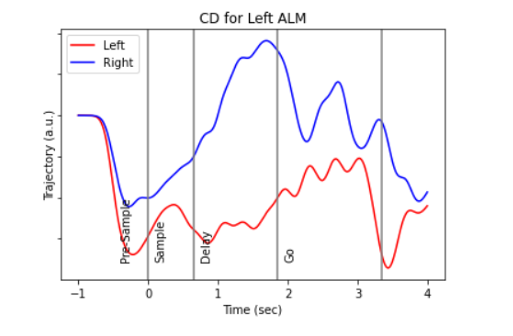
Other functions/Methods/Techniques???

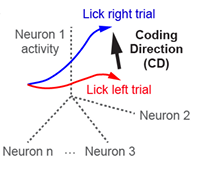
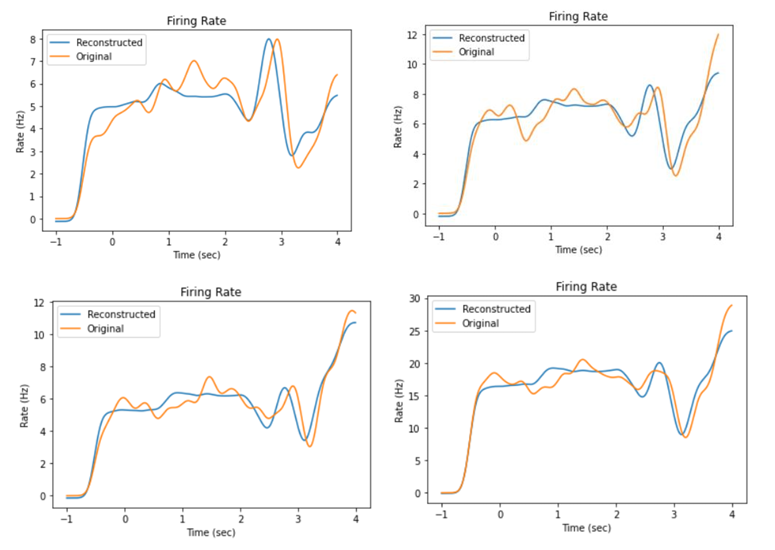
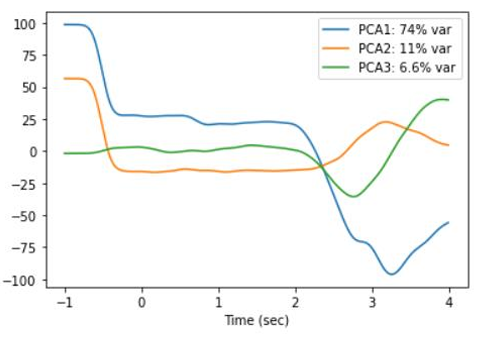
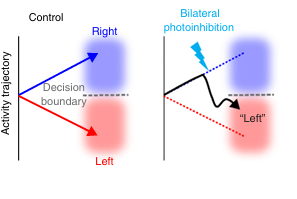




**Dimensionality Reduction** for Complex structure

Coding Direction and PCA ( Encoding/ Decoding of patterns (the exact detailed explanation in Internship form in gwdg email)



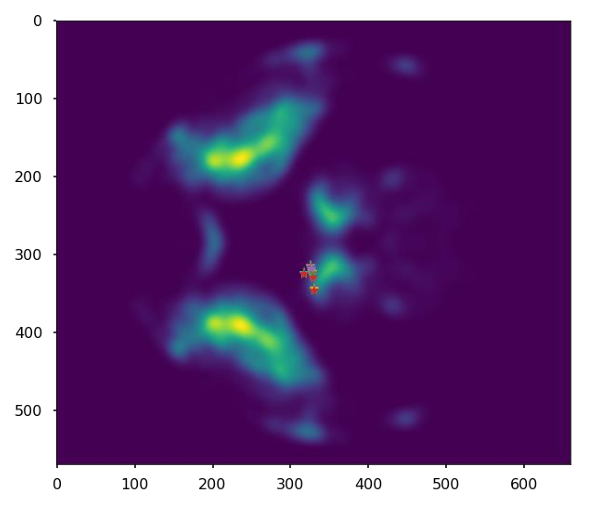
(Factor Analysis??)

* **Flow of the Pipeline/architecture/work flow:**

Try to come up with some flow chart/Diagram explaining work flow.

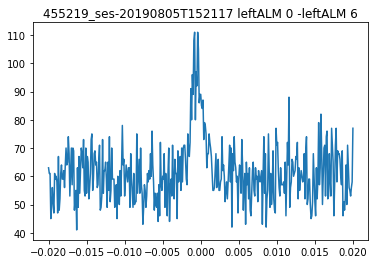
**Multimodal Analysis**

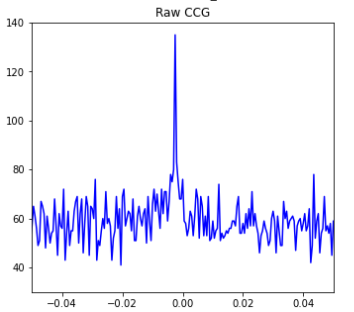
* **CCG**
  + **Jitter**
  + **Jitter Corrected**
  + **Peak filtering**
    - **Integrating for area (peakwidth)**
    - **Peak/std >= n-fold**
* **Fluorescence Analysis/Overlap**
  + **Projection Zones** (ALM projection zones for Thalamic neurons\_Chen et)
* **CCF-CCG Correlation**
* **Autocorrelograms (State and Answer the hypothesis from sticky notes)**
  + **Tau** (intrinsic time scale and inverse exponential curve fit)
  + **Peaks Criteria**
  + **Positive Tau** (Use Dynamic second peak And Try to fir Tau for this)
  + **Decaying Curves only**
* **P-values** 
  + **Test Statistics:** Spike Coincidence and Jitter CCG peaks(p<=0.05)

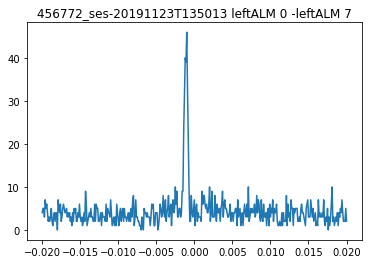
****

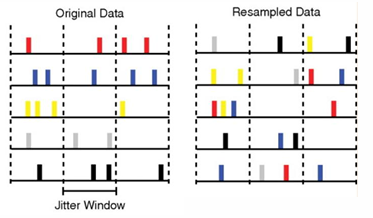
**RE-RUN THE PLOTS WITH PROPER AXIS TITLE**

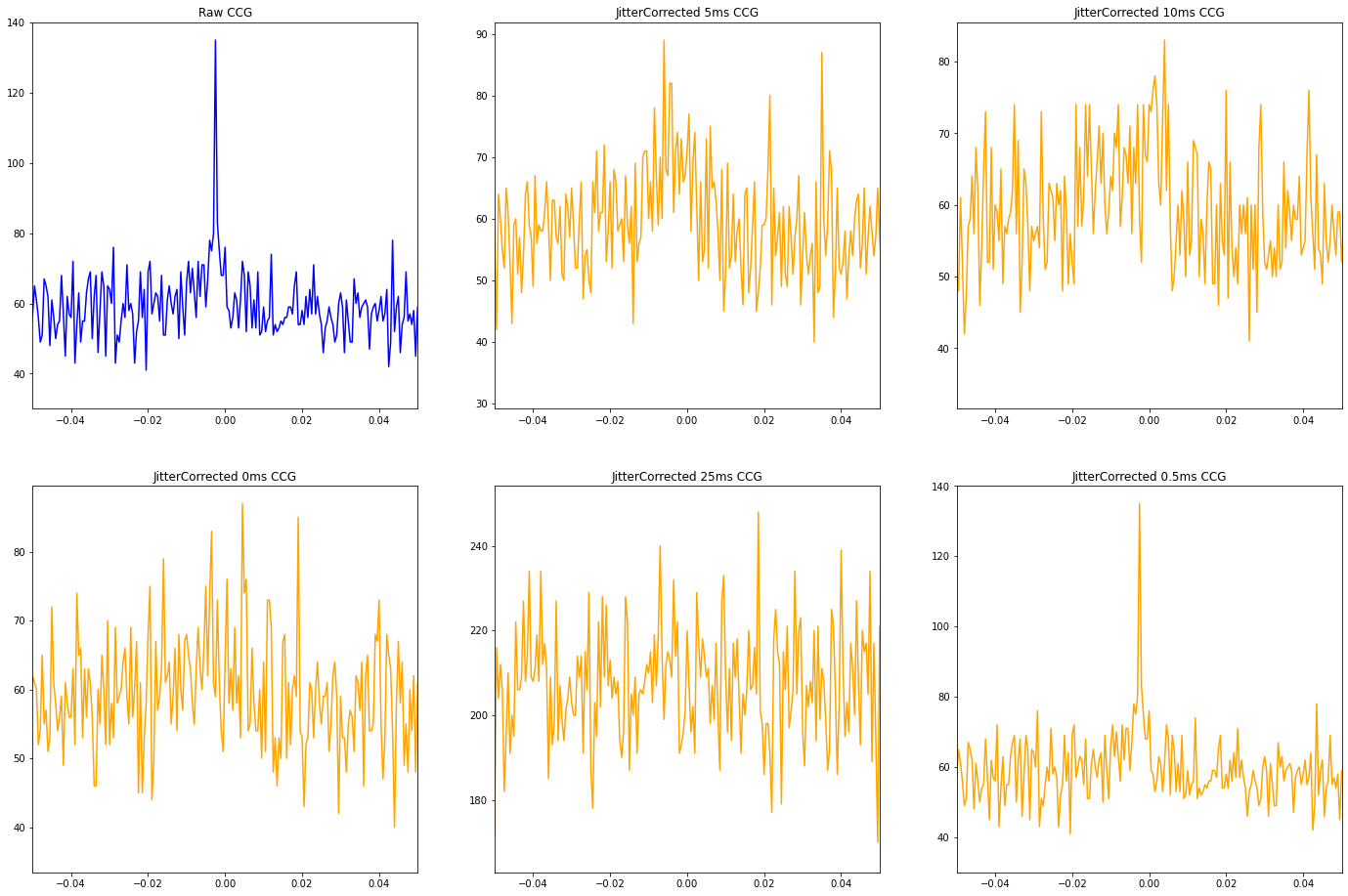
**VERIFY THE JITTER CODE AND PLOTS/RE-RUN THE JITTER CODE FOR PLOTS**

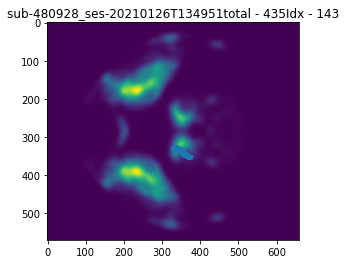
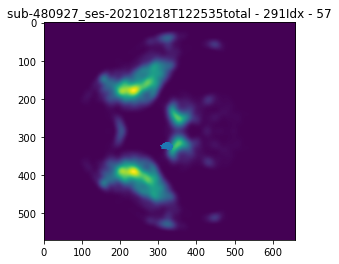
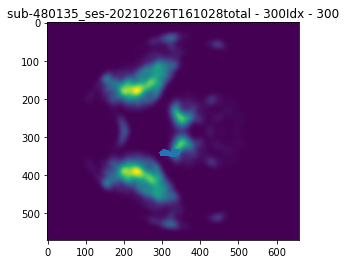
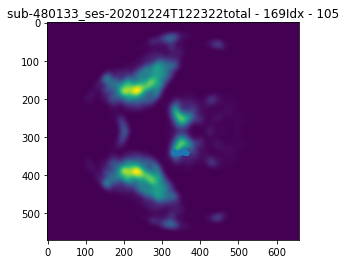
****

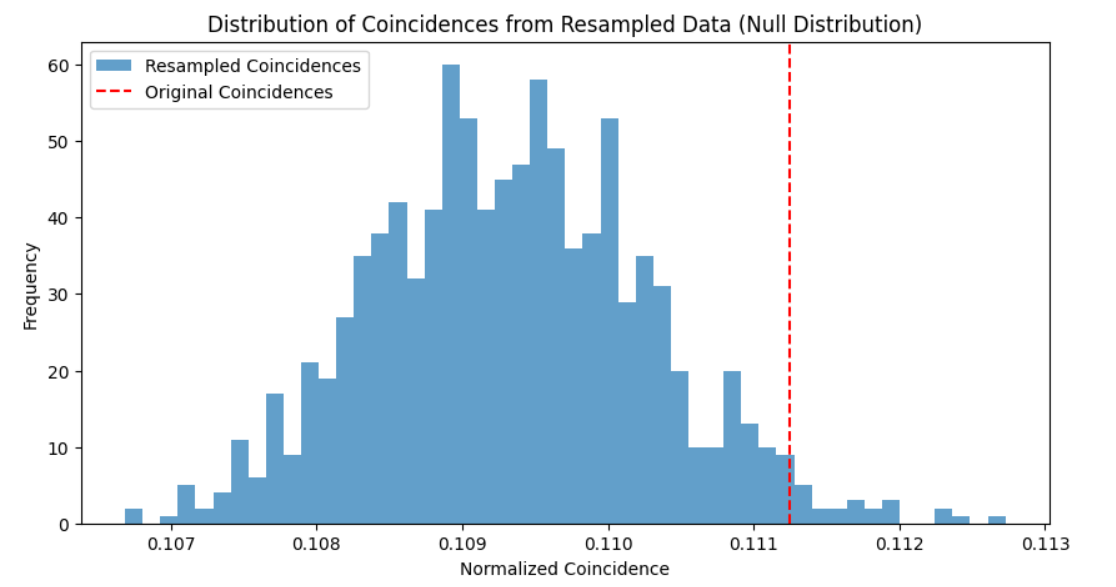
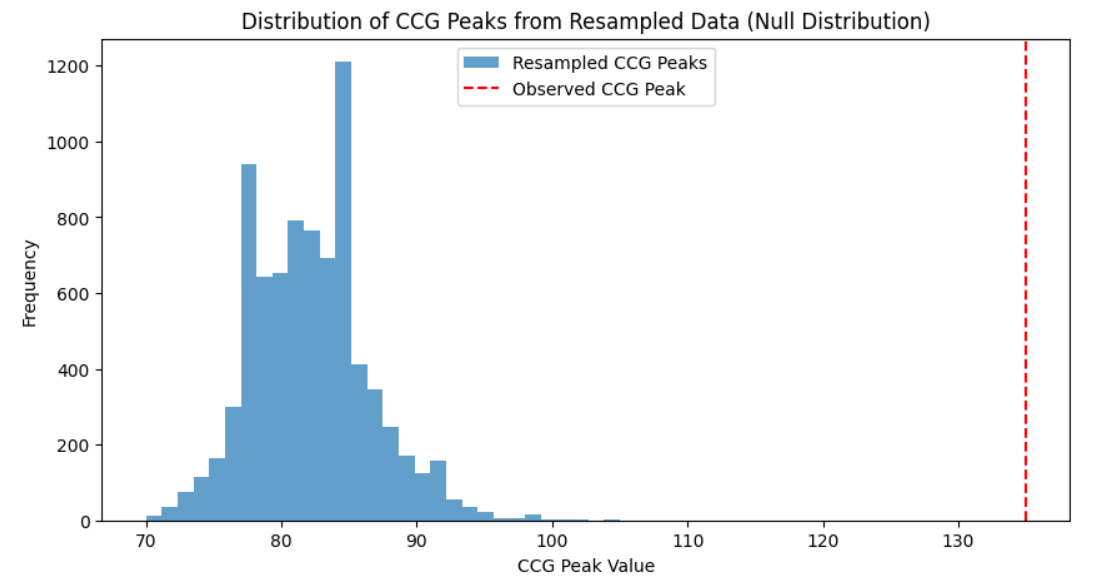
****

****

****

****

****



**Results and Discussion:**

Key Findings:

Contribution to Data/Data Analysis/Field

**Future Work/Direction (& Limitation)**

Epoch Analysis

Video Data/Medulla-Tongue movement

Multiregional Loops

>=3 Brain regional Interactions

**Conclusion**

**References**

In recent years, neuroscience has seen remarkable progress, particularly with the advent of advanced technologies that allow for extensive recording of neuronal activities, either sequentially or simultaneously. Despite these advancements, directly observing functional interactions at the cellular level throughout the brain remains a formidable challenge. This is largely due to the complexity of recording neural activities across numerous regions simultaneously, which limits our ability to fully capture the dynamics of brain networks involved in complex behaviors like decision-making and motor planning (Siegle et al., 2021).

Technological innovations in multi-electrode and optical recording have made it possible to monitor neuronal activities in both cortical and, in some cases, deeper structures. As these recording technologies continue to advance, they promise to exponentially increase the number of neurons that can be recorded simultaneously (Cunningham & Yu, 2014). This development is essential for understanding brain functions at a network level, where different regions work together to support cognition and behavior. Studies such as those by Steinmetz et al. (2019) and Stringer et al. (2019) have demonstrated the power of these technologies in capturing brainwide activity patterns, offering insights into how spontaneous and task-driven activities are encoded across multiple regions.

Previously, analytical techniques like cross-correlation and dimensionality reduction (e.g., Principal Component Analysis) have been widely employed to study neural dynamics. These techniques have proven effective for single-region data but fall short when applied to multi-region recordings due to limitations in capturing the full scope of interactions between distinct brain areas. Traditionally, single-probe recordings could only capture local neural dynamics, limiting researchers’ ability to analyze how multiple regions interact in coordinated and synchronous patterns (Svoboda & Li, 2018). Multiprobe recordings, however, enable simultaneous monitoring across multiple brain areas, enhancing spatial and temporal resolution and allowing researchers to study coordinated neuronal activities with greater detail and fidelity. This capability provides a comprehensive perspective on brain dynamics, which is essential for understanding complex neural functions and addressing research questions about inter-regional neural interactions (Chen et al., 2023).

The primary objective of this thesis is to unravel the mechanisms underlying movement planning processes involved in memory-guided tasks, using advanced statistical and machine learning methods. Specifically, the research focuses on understanding the contributions of individual neurons and population-level activity, shedding light on interactions within the Anteromedial Prefrontal Cortex (ALM) and its involvement in decision-making. Furthermore, this study seeks to examine how other brain regions, such as the Thalamus, interact with the ALM during decision-making processes. By considering additional factors like anatomical location and intrinsic timescales, this research aims to advance our comprehension of how these areas contribute to decision-related neural computations.

**Relevant Work and Knowledge Gap**

Advances in recording technologies, such as multiprobe and high-density electrode arrays, have enabled brain-wide neural recordings with improved spatial and temporal resolution, allowing researchers to capture distributed coding of choice and action across multiple brain regions (Steinmetz et al., 2019; Stringer et al., 2019). These large-scale, multi-region recordings have highlighted the importance of studying neural activity across interconnected areas to understand complex behaviors like movement planning. Despite these advancements, analyzing high-dimensional data remains challenging, prompting the use of dimensionality reduction techniques, such as Principal Component Analysis (PCA), to distill essential patterns from large-scale recordings (Cunningham & Yu, 2014). Although widely used, PCA alone may not fully capture complex, multi-region interactions, leading researchers to explore methods like Canonical Correlation Analysis (CCA), which examines shared neural variance between regions, thereby enhancing insights into inter-regional connectivity (Pehlevan et al., 2023).

The concept of intrinsic timescales has also emerged as a key factor in understanding how different brain regions process information over time, with research showing a hierarchical organization across cortical areas. Higher-order regions exhibit longer timescales, which support sustained cognitive processes such as decision-making and memory (Murray et al., 2014). This temporal hierarchy provides a framework for studying regions like the ALM and Thalamus, which may require different timescales for coordinated activity during tasks that involve memory-guided movement planning. Further studies on intrinsic timescales highlight the role of regions with longer timescales in sustaining task engagement, which is particularly relevant to interpreting temporal integration differences in ALM-Thalamus interactions during decision-related tasks (Hasson et al., 2015; Chaudhuri et al., 2015).

To understand inter-regional interactions, researchers have also investigated neural signal transmission across brain regions. Cross-Correlogram analysis, for example, evaluates temporal correlations and functional connectivity between neural populations, potentially revealing whether activity in one area, such as the ALM, is temporally correlated with activity in another, such as the Thalamus (Amarasingham et al., 2009). This method sheds light on the directional flow of signals, which is crucial for understanding brain network dynamics. Granger causality analysis further complements this approach by identifying whether neural activity in one region causally drives activity in another, making it a valuable tool for examining directed connectivity between the ALM and Thalamus (Gokcen et al., 2023). Together, these methods advance our understanding of the functional relationships that underlie complex behaviors**.**

**Knowledge Gap**

While previous studies have demonstrated the importance of multi-region recordings and dimensionality reduction, a gap remains in understanding the specific interactions between regions like the ALM and Thalamus during memory-guided movement tasks. Many existing studies lack the simultaneous recordings necessary to capture nuanced inter-regional dynamics involved in decision-making. Moreover, techniques such as PCA and CCA have been individually applied but seldom combined with connectivity-focused methods to study ALM-Thalamus interactions in depth.

This research addresses these gaps by leveraging simultaneous recordings and an ensemble approach—integrating Cross-Correlogram analysis, jitter analysis, p-value computation, and fluorescence signal analysis—to robustly quantify ALM-Thalamus connectivity. This combination enhances the reliability of our findings, offering a more comprehensive understanding of the neural basis of movement planning in mice.

### ****Data Source and Experimental Setup****

This study utilizes pre-recorded neural and behavioral data from 28 mice, collected by the Janelia Research Campus and described by Guo et al. (2014). The subjects included 25 VGAT-ChR2 EYFP mice (Jackson Laboratory, JAX #014548), one C57BL/6J (JAX #000664), one Sst IRES-Cre (JAX #013044) crossed with reporter mouse Ai32 (Rosa26-LSL-ChR2-EYFP, JAX #012569), and one Emx1-Cre (JAX #005628) crossed with R26-LNL-GtACR1 Fred-Kv2.1 reporter mouse (JAX #033089). A summary of recordings per mouse is presented in Table 1.

All animal procedures were conducted in compliance with protocols approved by the Janelia Research Campus Institutional Animal Care and Use Committee. Mice were maintained on a 12:12 reverse light-dark cycle, with behavioral tasks conducted during the dark phase. Each recording session lasted 1 to 2 hours, during which mice received their daily water intake (0.3–1.5 mL) to encourage task engagement. All surgical procedures, including titanium headpost implantation and craniotomies, were conducted aseptically. Analgesia (buprenorphine, 0.1 mg/kg, IP injection; Ketoprofen, 5 mg/kg, SC injection) was provided perioperatively, and mice were allowed to recover for at least three days with free access to water before starting water restriction.

The data collection setup included head-restrained, awake mice performing a memory-guided, auditory delayed response task (Inagaki et al., 2018). During the sample epoch, instruction tones of either 3 kHz or 12 kHz were played three times (150 ms each, 100 ms intervals). A delay period of 1.2 s followed, concluding with a ‘Go’ cue (6 kHz carrier frequency, 360 Hz modulation, 100 ms duration), prompting the mice to indicate the instruction by licking one of two ports. Correct responses were rewarded with a small water reward (0.1–0.2 µL), while incorrect responses triggered a timeout (1–3 s). To maintain task quality, trials with early licking or no response were excluded from analysis.

High-fidelity neural recordings were obtained through craniotomies (1–1.5 mm diameter), with three to four craniotomies per subject. Neuropixels probes were strategically inserted at various depths, targeting regions from the anterior lateral motor cortex to the medulla, providing a multi-region perspective on neural dynamics. Data was recorded in time-series formats including BehavioralEvents, BehavioralTimeSeries, Unit, and ElectrodeGroup, with a particular focus on spike times, which provide key insights into the neural dynamics of decision-making and motor planning behaviors.

Additionally, orofacial movements were tracked using two high-speed CMOS cameras (CM3-U3-13Y3M, FLIR) positioned for side and bottom views (Mathis et al., 2018). Videos captured at 300 Hz under IR illumination (940 nm LED) were analyzed with DeepLabCut to track the tongue, jaw, and nose movements, aiding in the analysis of task-related orofacial dynamics (Mathis et al., 2018).

**Data Engineering**

This section outlines the steps taken to preprocess both the Trial data and Unit Data, including merging, cleaning, aligning, and preparing [ADD MORE TECHNIQUES] these datasets for subsequent

* **Retrieval and Curation:**

The datasets were stored in DANDI repository where to retrieve data we need to use either the Python Command Line Interface (CLI) Client or the DANDI Web application. Specific subject data can be identified and downloaded by querying the DANDI web platform or using CLI commands to streamline retrieval. After determining the relevant subject ID, individual file or session links can be used to access specific datasets through the DANDI CLI.

Each session's data is stored in the Neurodata Without Borders (NWB) format, an open standard designed for organizing large-scale neurophysiological data. Each .nwb file typically contains comprehensive time-series and video data, amounting to hundreds of gigabytes per session, which includes electrophysiological recordings, behavioral events, and high-speed video captures. This format facilitates the integration of data from different modalities and time scales, allowing for standardized storage and efficient data access.

Following data retrieval, the next step involves curating and organizing the data for each mouse. The curation process consolidates individual session data into a main Data Table, preserving the details of each session while providing a structured overview at the subject level. This structured organization aids in subsequent analyses, enabling consistent access to session-specific details while maintaining a comprehensive dataset framework.

* **Pre-procession/Cleaning:**

**From Version3 and use following terms to make it look professional.**

Data Transformation: Defining/segmenting in epochs, Spike binning (kernel smoothing)

**Consistent Referencing**: Ensure consistent referencing for each neural unit, electrode, and anatomical region. Keeping track of these references across sessions is important for multi-region analyses. This can be facilitated by assigning unique identifiers to electrodes and neurons and using metadata files or lookup tables.

**Anatomical Filtering**: Focus on data from specific brain regions, such as ALM and Thalamus, by filtering units or channels based on their anatomical metadata. This is typically specified in tables or metadata files in the NWB dataset

**Temporal Alignment:** Align the start and end times of each epoch (e.g., stimulus, delay, and go) to ensure temporal consistency. Any session where epochs are not sequential or contain time gaps should be flagged for potential removal or correction.

**Trial Exclusion Criteria**: Trials with abnormal durations (outliers) or where epoch timings are inconsistent (e.g., delay periods lasting outside of the expected 1–1.3 seconds) should be excluded. Establishing a maximum trial duration (e.g., 5.3 seconds in our case) helps maintain uniformity in trial analysis.

**Epoch Outliers**: distribution of duration(check very high/low duration nd remove them)

**Spike Outliers:** Spike with low firing rate

Additional Queries as per Requirements and further analysis??

**Manual Review:** Visualize key data characteristics (e.g., spike rasters, firing rate histograms, epoch distributions) to manually spot-check for any anomalies that automated checks might miss. Reviewing a subset of trials per subject can also help confirm that processing steps are being consistently applied**.(THIS IS SOMETHING I DID)**

**Methodology**

**An outline/process of this entire project in visual representation via flow-chat/Figure will help to understand it better. [FIGURE]**

**Spike Statistics:**

**Firing Rate:**

In this section, we focus on the analysis of spike statistics, specifically employing firing rate calculations and Gaussian kernel smoothing to represent neural activity in a meaningful way. Neuronal spikes, or action potentials, serve as the fundamental unit of information in the brain, which encodes information through the timing and frequency of these events (Abbott, 2001). Spikes can be grouped into spike trains or collective patterns of activity that reveal the underlying structure and function of neural circuits. Understanding these patterns through spike statistics is critical for decoding the neural representations and processing mechanisms within brain regions such as the Anteromedial Prefrontal Cortex (ALM) and Thalamus.

Fig From book here

**Firing Rate:**

One of the key metrics used in spike train analysis is the **firing rate**. The firing rate measures the frequency of action potentials fired by a neuron over a defined period, providing a quantitative basis for interpreting neural activity patterns. Formally, the firing rate (FR) is calculated as:

FR=S/T​

where:

* **FR** is the firing rate (spikes per second),
* **S** is the total number of spikes observed,
* **T** is the time interval over which the spikes are counted.

By determining the firing rate, we can observe trends in neuronal activity across different behavioral epochs, such as the sample, delay, and response phases in memory-guided tasks. In this study, histograms and window functions are applied to visualize the firing rates, which allows for intuitive interpretation of how neuronal populations respond to task-relevant stimuli and periods.

Gaussian kernel Smoothing:

To obtain a continuous and smooth representation of the firing rate, we apply **Gaussian kernel smoothing**. This technique involves convolving the discrete spike train with a Gaussian kernel, providing a smoothed approximation of the underlying firing pattern. Gaussian kernel smoothing is especially valuable in studies with high-frequency spike data, as it reduces noise and enhances the interpretability of neural activity over time.

In our analysis, we applied a Gaussian kernel with a sigma (σ) of 0.1[CITE FIGURE], a parameter chosen to balance temporal resolution with smoothness. The kernel width effectively determines how much the firing rate at each point is influenced by surrounding time points, with a smaller sigma providing higher resolution and a larger sigma yielding a broader, smoother curve. By adjusting sigma, we optimize the representation of activity trends without losing critical temporal details.

The smoothed firing rate offers a continuous estimate that captures the dynamics of neuronal responses over time, making it easier to interpret changes in firing patterns across task conditions. This approach is particularly useful in identifying key periods of neural activity modulation, such as increases in firing rate during response epochs, that may correlate with decision-making or motor planning processes.

providing a clear representation of neural activity fluctuations over task epochs. These visualizations allow us to

 Compare activity across different brain regions (e.g., ALM vs. Thalamus) and assess how firing rates change in response to specific task phases.

 Identify patterns in neural responses, such as sustained activity during delay periods, which may be indicative of memory encoding or anticipation.

 Facilitate comparisons across trials, aiding in the identification of consistent activity patterns tied to the behavioral task.

**Population Analysis:**

Coding Direction and PCA:

[FOR PCA THERE WERE MORE PAPERS]

In this study, population analysis is used to investigate collective behavior of neurons during decision-making and motor planning tasks. By examining activity patterns across populations of neurons, we aim to capture higher-level neural dynamics that contribute to task-specific outcomes. This analysis employs two primary techniques: **Coding Direction** and **Principal Component Analysis (PCA)**, both of which are effective for extracting meaningful patterns from complex, high-dimensional neural data.

**3.8.1 Coding Direction Analysis**

The Coding Direction (CD) method is used to characterize and distinguish neural responses associated with specific behavioral outcomes, such as leftward or rightward licks, in the context of a memory-guided task. In this analysis, a set of orthogonal directions in the multi-dimensional activity space is defined, each represented by an n×1 vector, where n is the number of neurons. These vectors are designed to maximize the separation between response vectors for lick-left and lick-right trials at task-relevant time points within this n-dimensional activity space (Chen, 2023).

To compute the Coding Direction, we performed the following steps:

[REVISE THE MATHS FORMULA]

1. **Mean Spike Count Calculation**: For each trial type (lick-left and lick-right), we calculated the mean spike count vector separately for each neuron. This provides two distinct mean response vectors: x⃗lick-left\vec{x}\_{\text{lick-left}}xlick-left​ and x⃗lick-right\vec{x}\_{\text{lick-right}}xlick-right​.
2. **Direction Calculation**: At each time point, we computed the difference between these mean vectors, yielding the vector w⃗t=x⃗lick-left−x⃗lick-right\vec{w}\_t = \vec{x}\_{\text{lick-left}} - \vec{x}\_{\text{lick-right}}wt​=xlick-left​−xlick-right​. This vector w⃗t\vec{w}\_twt​ represents the Coding Direction that maximally separates the two trial types.
3. **Averaging Across Epochs**: To obtain a robust representation of the Coding Direction over the task period, we averaged w⃗t\vec{w}\_twt​ across specified epochs (e.g., delay and go periods).
4. **Projection onto the Coding Direction**: Finally, we projected the activity of each trial onto the Coding Direction vector. This projection facilitated a quantitative assessment of how well the neural population activity aligned with the task demands, specifically the discrimination between lick-left and lick-right trials (Chen, 2023).

[INSERT FIGURE AND CITE AND DESCRIBE FIGURE]

[DECISION BOUDARY ALSO]

Through this process, we observed that the population activity in regions like the anterior lateral motor cortex (ALM) and the Thalamus could be visualized in an activity space where trajectories reflect decision-relevant dynamics. For instance, in the ALM, the Coding Direction showed distinct clustering of trajectories corresponding to lick-left and lick-right trials, especially during the late delay epoch (last 0.6 seconds), allowing us to identify clear boundaries between the two trial types. By applying this approach to both the ALM and Thalamus, we found that Coding Direction analysis revealed distinct neural representations for each response type across critical task phases, particularly in the late delay and early go periods, making it ideal for examining selectivity.

**3.8.2 Principal Component Analysis (PCA)**

Principal Component Analysis (PCA) is a widely used dimensionality reduction technique that identifies orthogonal directions (principal components) in the data that capture the greatest variance. In this study, PCA was applied to simplify the high-dimensional neural activity data, representing complex firing patterns through a reduced set of key components. By capturing the primary variance in the neural firing rates, PCA allows us to represent underlying trends in population activity, aiding in the extraction of patterns relevant to memory-guided tasks.

[INSERT FIGURE AND CITE AND DESCRIBE FIGURE]

The following steps were conducted to apply PCA to the neural activity data:

1. **Variance Maximization**: PCA identifies the directions that maximize the variance in the data. Each principal component (PC) represents an axis in the lower-dimensional space that captures significant variance, allowing us to simplify complex population activity into a few key components (Cunningham & Yu, 2014).
2. **Low-Dimensional Projection**: Neural activity data for each trial was projected onto the first few principal components, reducing dimensionality while retaining the most informative variance. This projection highlights patterns in firing rates across trials, revealing population-level structures within the data.
3. **Interpretation of Principal Components**: Each principal component was analyzed to understand its contribution to the overall variance. In this study, the initial appearance of neuron firing rates was seemingly random and high-dimensional. However, PCA revealed that the underlying structure could be distilled into a few principal components, with these components effectively capturing relevant trends across different behavioral conditions.

[REMOVE THE REDUNDANCY AND CITE FIGURE/PAPER]

**Interpretation**

While PCA is highly effective in reducing dimensionality, it is important to note that the variance captured by PCA includes all types of variability, including both task-relevant and spontaneous fluctuations in firing rates (Cunningham & Yu, 2014). This means that the low-dimensional space identified by PCA may include variability unrelated to task events, which could introduce noise into the interpretation of neural dynamics. Despite this limitation, PCA remains a valuable tool for summarizing neural population activity and identifying general patterns across trials.

**END NOTE**

* The combination of these two methods allowed us to map decision-related neural responses in regions such as the ALM and Thalamus, supporting the analysis of complex multi-dimensional activity patterns.

**Multimodal Analysis:**

**Cross correlation Analysis:**

**CCG:** Cross-Correlogram analysis reveals significant time-lagged correlations between the Anterior Lateral Motor Cortex (ALM) and Thalamus, indicating potential directional signal transmission during the memory-guided task.

[Graphs And MORE PAPERS CITE/FIGURE/JITTER]

**Jitter corrected CCG:**

[DESCRIPTION FOR JITTER][PLOTS]

**AutoCorrelograms and Intrinsic timescale (tau):**

[PLOTS AND DESCRIPTION]

**P-values:**

**Fluorescence analysis:** Fluorescence signal analysis of projection zones confirms targeted input from the Anterior Lateral Motor Cortex (ALM) to the Thalamus, supporting its involvement in the multi-regional circuitry of the delayed-response task.