Multi-sensory integration is a fundamental problem for any embodied cognitive system – both biological and artificial. We pursue a network diffusion approach to model the flow of evoked activity, initiated by stimuli at primary sensory regions. In particular, we apply the Asynchronous Linear Threshold (ALT) diffusion model on the mesoscale cortical connectome of the mouse. The ALT model captures how evoked activity at a given cortical region ripples through the rest of the cortex. Our results show that a small number of regions (the Claustrum being at the top of the list) integrate almost all sensory information paths, suggesting that the cortex relies on an “hourglass architecture” to integrate and compress sensory information before utilizing that lower-dimensionality representation in association areas and higher cognitive tasks.

Keywords: Multi-sensory integration; mouse connectome; network diffusion; hourglass architecture; Claustrum

Introduction

Consider a system that has several sensory inputs (e.g., visual, auditory, haptic) and that needs to encode, integrate and compress this multi-modal stream into a consistent perceptual state before it can respond. Further, the environment may be noisy and time-varying, and so any single sensory stream may not be sufficient for robust perception. Both biological and artificial intelligence systems must solve the problem of Multi-Sensory Integration (MSI), and thus this is a fundamental problem for any embodied cognitive agent.

In the brain, MSI takes place across widely distributed networks (Meredith, 2002; Stein, Stanford, Ramachandran, Perrault, & Rowland, 2009). The neural basis of how sensory streams are encoded, integrated and compressed is an increasingly active research area in neuroscience (Meijer, Mertens, Pennartz, Olcese, & Lansink, 2019).

In artificial agents, and in deep learning in particular, we may not be sufficient for robust perception. Both biological and artificial intelligence systems must solve the problem of Multi-Sensory Integration (MSI), and thus this is a fundamental problem for any embodied cognitive agent.

For a given activated source node, we calculate the activation time of all other nodes that participate in the cascade. We then construct a Directed Acyclic Graph (DAG), dubbed Activation-DAG (A-DAG), based on the nodes’ activation times for each type of sensory stimulation. We analyze the result-
ing A-DAGs to identify salient architectural properties of the mouse cortex that enable MSI.

Our results show that a small number of cortical ROIs (the Claustrum being at the top of the list) integrate almost all sensory information streams. This suggests that the cortex relies on an hourglass architecture (see Fig. 1). An hourglass architecture has many input modules at one end of the architecture (one or more for each sensory modality), many output modules at the other end (one or more for each cognitive or motor task) and only a small number of core modules at the waist of the architecture (Friedlander, Mayo, Tlusty, & Alon, 2015; Sabrin & Dovrolis, 2017; Sabrin, Wei, van den Heuvel, & Dovrolis, 2019). The objective of those core modules is to compute representations of the input space that are both efficient (i.e., of low dimensionality compared to the inputs) and accurate (i.e., they can capture almost all of the input variance). Further, these core modules tend to remain invariant, even when there are changes in the input and/or output modules (Siyari, Dilkina, & Dovrolis, 2018).

To validate the ALT model, we use Voltage Sensitive Dye (VSD) imaging data (Mohajerani et al., 2013). VSD data was collected from mice while different sensory stimuli were introduced to evoke responses in the visual, somatosensory (upper limb, lower limb and whisker), and auditory cortices. This analysis corroborated the predictive power of ALT in modeling diffusion of activity from specific source ROIs.

![Figure 1: A hypothetical hierarchy with feedforward, feedback and lateral connections between modules. Input information is provided at sensory-specific modules, while the high-level cognitive tasks are performed by task-specific modules at the other end.](image)

**Summary of methods and results**

The Allen Mouse Brain Atlas has 213 ROIs (Oh et al., 2014). We consider a network of 67 ROIs that reside in the right hemisphere of the isocortex, olfactory areas, hippocampal formation and cortical subplate. The edge weights we consider are referred to as “connection density” (Oh et al., 2014). This metric is roughly proportional to the average number of axons from the source ROI that target neurons connect to. We filter the anatomical edges with p-value higher than 5%. We denote the resulting 67-node weighted network by \( N_c \); its density is 13%. Edge lengths are calculated by the Euclidean distance between ROI centroids.

We model the diffusion of evoked neural activity using an Asynchronous Linear Threshold model. The incoming neighbors of a node (or set of nodes) are denoted by \( N_{in}(n) \). Initially, the binary state of every node is set to 0, except a given primary sensory ROI whose state is set to 1. Subsequently, the state of each node \( n_i \) is updated asynchronously based on the state of all its incoming neighbors’ \( N_{in}(n_i) \), as follows:

\[
s_i(t) = 1 \text{ if } \sum_{j \in N_{in}(i)} w_{ji} s_j(t - t_{ji}) > \thetaanumber{1}\]

where \( \theta \) represents the activation threshold, \( t_{ji} \) the propagation delay from node \( j \) to \( i \) and \( w_{ji} \) is the connection density of the connection \( j \rightarrow i \).

The sensory ROIs we consider as source nodes are: AUDp (primary auditory), VISP (primary visual), GU (gustatory areas), SSP-n (somatosensory – nose), SSP-bfd (somatosensory – barrel field), SSP-ll (somatosensory – lower limb), SSP-ul (somatosensory – upper limb), SSP-m (somatosensory – mouth), SSP-tr (somatosensory – trunk), and MOB (main olfactory bulb).

For each of these ten source nodes, we calculate the activation time of all nodes that participate in the ALT cascade. We construct a Directed Acyclic Graph (DAG), hereafter dubbed Activation DAG (A-DAG), based on the nodes’ activation time. The A-DAG includes an edge \( n_i \rightarrow n_j \) if and only if \( n_j \in N_{out}(n_i) \) and \( t_j > t_i + t_{ij} \), i.e., only those nodes that contributed to the activation of \( n_i \) should point to \( n_j \) in the activation DAG.

We have investigated the size of the cascade (number of activated nodes in the cascade) for different values of \( \theta \). Interestingly, either all nodes participate in the cascade or almost none of them does. Thus, we set \( \theta \) to the highest value that causes a complete cascade (\( \theta = 0.46 \) for MOB, which is the only sensory ROI outside the isocortex, and \( \theta = 0.98 \) for all other sources). Complete cortical cascades of sensory activity have been experimentally observed (Mohajerani et al., 2013).

ALT is used to capture the activation propagation from a local perturbation to the whole cortical network. This propagation captures the first time the impact of a focal perturbation reaches each ROI. We do not expect ALT to model any subsequent feedback or sustained oscillations between ROIs.

We analyze the resulting A-DAGs based on the activation paths that originate from sensory nodes (the source of each A-DAG) and terminate at each sink node of that A-DAG. The Path Centrality (PC) of each node is defined as the fraction of source-to-target paths traversing the node across all A-DAGs.

We then compute the \( \tau_{core} \) of the network i.e., the minimum set of nodes that covers at least \( \tau \% \) of all source-to-target paths across all activation DAGs (Sabrin & Dovrolis, 2017). This NP-Hard problem is approximated by a greedy
heuristic in which the node with maximum PC is removed from
the network in each iteration, joining the $\tau$−core set. The PC
of the remaining nodes is updated after each iteration (Sabrin
& Dovrolis, 2017).

Interestingly, when we compute the $\tau$−core for $\tau=90%$
across all A-DAGs, we find only nine nodes. These core nodes
are listed in table 1. With each core node, we also list the fra-
tion of source-target paths it covers ("coverage") when it joins
the core as well as its PC rank. The table identifies CLA, PTLp
and AUDv as the three most important ROIs for MSI – these
three ROIs cover more than 50% of the source-target paths in
the ten A-DAGs.

The claustrum (CLA) is known for its anatomical unique-
ness and its enigmatic function (Crick & Koch, 2005; Van
Horn, 2019) – Francis Crick had hypothesized that the
Claustrum plays a central role in the emergence of conscious-
ness. In an intriguing experiment, Koubelissi et al. (Koubelissi,
Bartolomei, Beltagy, & Picard, 2014) found that by delivering
electric pulses to CLA in a human subject, she was immedi-
amately driven to a “frozen state” in which she could not con-tinue
a reading task – discontinuing the stimulation immediately re-
sumed normal behavior.

The posterior parietal associative area (PTLp) has strong
and direct connectivity to primary sensory ROIs and projec-
tions to motor areas, and it has been previously identified as
a "hub" in the responses from a variety of sensory stimulation
responses (Lim et al., 2012; Mohajerani et al., 2013; Nikbakht,
Tafreshiha, Zoccolan, & Diamond, 2018).

To examine the robustness of this hourglass observation,
we have examined whether the $\tau$−core stays small, and
whether it consists of the same nodes, if we randomize the
underlying connectome. Randomizing the connection weights
and/or lengths, but preserving their distribution, does not have
a significant effect on the size or membership of the $\tau$−core
– the hourglass waist is not affected. If we randomize the

topology of the connectome however, by swapping edges in a
degree-preserving manner, the $\tau$−core doubles in size and
it changes significantly in membership. We conclude that it is
the architecture of the connectome, rather than the weights or
lengths of the connections, that are the primary reason behind
the hourglass structure of MSI.

We have used Voltage Sensitive Dye (VSD) imaging data
(Mohajerani et al., 2013) to examine the accuracy of the ALT
modeling predictions. VSD imaging enables neural population
activity monitoring over large cortical areas and in temporal
resolution of few milliseconds. Each VSD video corresponds
to a single sensory stimulation experiment on a mouse. The
resulting videos have 108 frames at a temporal resolution of
6.67ms. Each frame consists of 128 × 128 pixels at the spa-
tial resolution of 50µm/pixel. We compare ALT to VSD im-
ages for five animals and five sensory stimulation types: visual
(VISp), auditory (AUDp), whisker (SSp-bfd), forelimb (SSp-ul)
and hindlimb (SSp-ll).

To assign an activation frame for each ROI in the VSD data,
we performed the following steps: 1) Find the peak signal for
each VSD pixel (the frame at which the peak activity is ob-
served). 2) Register the Allen Atlas ROIs to each animal’s
native space. 3) Find the activation frame for each ROI based
on the activation frame of the majority of that ROI’s pixels.

We then compare the temporal ordering of activated ROIs
between VSD experiments and ALT modeling. There are
three cases: 1) **Temporal agreement** comprises of all ROI
pairs for which ALT predicts correctly the experimental activa-
tion order. 2) **Insufficient temporal resolution** comprises
of all ROI pairs that appear to get activated in the same VSD
frame. 3) **Temporal disagreement** comprises of all ROI pairs
for which ALT does not predict correctly the experimental ac-
tivation order. The results are summarized in Fig. 3 – note
that the frequency of "temporal agreement" varies between
50-75% across five sensory stimulations and five animals per
stimulation – while the frequency of "temporal disagreement"
varys between 0-20%.

**Brief discussion**

The mechanism through which the brain integrates differ-
ent sensory inputs and reaches a stable unified perception
has been under investigation at different spatial and tempo-
ral scales, with classical works focusing on subcortical ROIs,
such as the Superior Colliculus (Wallace & Stein, 1997; Stein

---

**Table 1: Core nodes across the ten A-DAGs ($\tau = 90\%$).**

<table>
<thead>
<tr>
<th>Node</th>
<th>Description</th>
<th>Coverage</th>
<th>Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>CLA</td>
<td>Claustrum</td>
<td>24</td>
<td>1</td>
</tr>
<tr>
<td>PTLp</td>
<td>Posterior parietal assoc.</td>
<td>19</td>
<td>3</td>
</tr>
<tr>
<td>AUDv</td>
<td>Ventral auditory area</td>
<td>19</td>
<td>4</td>
</tr>
<tr>
<td>SSs</td>
<td>Supp somatosensory area</td>
<td>9</td>
<td>2</td>
</tr>
<tr>
<td>MOs</td>
<td>Secondary motor area</td>
<td>9</td>
<td>5</td>
</tr>
<tr>
<td>ACAd</td>
<td>Anterior cingulate - dorsal</td>
<td>6</td>
<td>16</td>
</tr>
<tr>
<td>VSt</td>
<td>Lateral visual area</td>
<td>5</td>
<td>19</td>
</tr>
<tr>
<td>ADB</td>
<td>Accessory olfactory bulb</td>
<td>4</td>
<td>7</td>
</tr>
<tr>
<td>ECT</td>
<td>Ectominal area</td>
<td>4</td>
<td>22</td>
</tr>
</tbody>
</table>

---

Figure 2: The first 10 activated nodes in the A-DAG of
the somatosensory-nose ROI (SSp-n), according to the ALT
model. Red edges are connectomic connections that partici-
pate in the activation cascade, while grey edges do not.
Figure 3: Percentage of ROI pairs that show temporal agreement (green), insufficient temporal resolution (blue), and temporal disagreement (red) for five different stimulation types (x-axis labels) and five different animals (different symbols).

et al., 2009) to more recent larger scale foci (Toker & Sommer, 2019; Worrell, Rumschlag, Betzel, Sporns, & Mišić, 2017). After compiling decades of MSI work, Meijer et al. concluded that MSI research has been overwhelmingly focused on behavioral-level integration of external stimuli and the associated micro-circuitry, and he proposed a shift of focus to “a systems neuroscience approach, with rodents as the prime model, to investigate how the neocortex combines sensory stimuli of different modalities” (Meijer et al., 2019). This is the line of inquiry we have followed here. We found that (a) information flow originating from primary sensory cortices is reasonably well predicted by a linear threshold-based network diffusion model, (b) a small number of cortical ROIs integrate and mediate almost all sensory pathways, and (c) these “core nodes” are anatomically close to the primary sensory ROIs in the cortex (not shown here due to space constraints).

These findings support the existence of an “hourglass architecture” in the integration of multisensory information. The benefit of an hourglass architecture is that it first reduces the dimensionality of the inputs to a much lower dimensionality latent representation at the “hourglass waist”, and second, it re-uses those compressed intermediate-level features in more than one higher-level tasks (Sabrin & Dovrolis, 2017).

References


Van Horn, J. D. (2019). What is old is new again: Investigating and analyzing the mysteries of the claustrum. Springer.
