Hypnagogia: A Pre-Registered Test for Offline Memory Replay in a Whole-Brain *Drosophila* Connectome Model

Aniketh Bandlamudi
University of Virginia
zpq3uh@virginia.edu me@aniketh.net

November 2026

Abstract

Sleep replay of learned activity is well described in the rodent hippocampus, but offline reactivation of a learned ensemble has never been observed in an insect. We asked whether a whole-brain leaky integrate-and-fire model of the *Drosophila* male central brain connectome (144,209 neurons, 21.9 million connections), built on the model of Shiu et al. [28], reactivates a learned Kenyon-cell ensemble during simulated sleep. Our reimplementation reproduced the published model on its own data (per-neuron rate correlation $r = 0.9997$). We then corrected two mushroom-body cells against direct measurements, adding no parameter: APL became a non-spiking, graded-release neuron and DPM became inhibitory. The APL correction alone moved odour coding from 52.6% of Kenyon cells at 23.0 Hz to 4.5% at 0.79 Hz, against 6 ± 5% measured in flies. An odour memory was encoded by dopamine-gated depression at Kenyon-cell-to-output-neuron synapses and tested against four pre-registered comparisons over 20 seeds. The test failed. The memory changed offline Kenyon-cell activity in all 20 seeds and reduced drive through the learned synapses by 32.5%, but it did not increase reactivation of the trained ensemble (learned minus unlearned weights: +0.016, 95% CI −0.011 to +0.044). An injected positive control showed that the published model cannot hold a discrete episode. Adding spike-frequency adaptation, a labelled deviation, produced episodes and a sparser code, yet the memory still contributed nothing (+0.040, 95% CI −0.010 to +0.090). The model's only plastic synapses sit downstream of the cells being scored, and the path back delivers under 1% of a threshold gap even at 50 Hz. Kenyon-cell replay is therefore a question this model class cannot answer positively; the negative is a finding about the model, not about the fly.

Contents

1 Introduction **3**

2 Methods **3**
2.1 Connectome 3
2.2 Neuron model 4
2.3 Engine check and validation 4
2.4 Background noise and operating point 5
2.5 Two corrections to the mushroom body 5
2.6 Plasticity and memory encoding 5
2.7 Sleep state 6
2.8 Replay statistics and pre-registered criteria 6
2.9 Positive control and labelled deviations 7

3 Results **7**
3.1 The engine reproduces the published model, and the network is bistable 7
3.2 Correcting APL restores a sparse odour code 8
3.3 The offline state is the antennal lobe running away 9
3.4 A memory is written at the synapse, but the readout cannot show it 9
3.5 The replay test fails at published parameters 10
3.6 Why there is no replay 11
3.7 A slow variable produces episodes, but still no replay 12

4 Discussion **13**

5 Limitations **15**

6 Conclusion **15**

Code and Data Availability **15**

Acknowledgments and Use of AI Tools **16**

1 Introduction

In the rat hippocampus, ensembles that fire together during waking fire together again during later sleep [19, 30]. During pauses in waking behaviour, place cells can also replay the order in which they fired, in reverse [10]. Offline reactivation of a learned ensemble, above chance and above structural controls, has become the standard operational test for replay.

The fly has the ingredients for the same question but no answer to it. Activating dorsal fan-shaped body (dFB) neurons promotes sleep [7, 25], and R5 ring neurons show network-generated up and down states at 0.5 to 1.5 Hz whose power rises with sleep need [26]. The mushroom body stores odour memories: Kenyon cells code odours sparsely, about $6 \pm 5\%$ of cells per odour [13, 29], and dopamine-gated depression at Kenyon-cell-to-output-neuron (KC→MBON) synapses writes the association [12]. Sleep supports memory consolidation in flies [4, 11], and dopaminergic neurons active during training are reactivated in post-learning sleep, as read out by a transcriptional reporter over hours [6]. Yet offline reactivation of a Kenyon-cell ensemble has never been observed in a real fly, and the fly has no described analogue of the hippocampal sharp-wave ripple that organises mammalian replay.

Whole-brain connectomes make it possible to ask the question in simulation. Shiu et al. [28] built a leaky integrate-and-fire (LIF) model of every neuron in the FlyWire connectome [9] and showed that it predicts sensorimotor pathways for taste. The male central nervous system (CNS) connectome v1.0 [3] now provides a second complete, proofread brain. A model built on it lets replay be tested as a controlled experiment: the same network with and without learning, with real and shuffled wiring, in sleep and in wake.

This paper reports that experiment, a project we call Hypnagogia. Its contributions are:

1. a Brian2 reimplementation of the Shiu et al. [28] model that reproduces the published output, ported to the male CNS connectome and validated there;
2. two corrections to mushroom-body cells that the published model mis-specifies, each taken from a direct measurement and adding no parameter, of which the APL correction is the largest single effect in the project;
3. a pre-registered replay test with four required comparisons and an unlearned-weights control, run over 20 seeds, which is negative;
4. a diagnosis of the negative: the model cannot hold an episode, the memory is expressed tonically rather than in episodes, and the only plastic synapses sit downstream of the cells being scored;
5. a demonstration that adding a slow variable produces episodes but still no replay.

Every number below is read from the project's result files. Departures from the published model are labelled as deviations wherever they appear, and any constant without a measurement behind it is scanned rather than chosen.

2 Methods

2.1 Connectome

The network is the male CNS connectome v1.0 [3], restricted to the brain. Of 211,577 annotated bodies, 165,122 are traced. Removing ventral nerve cord, enteric and unclassified bodies leaves 144,209 neurons and 21,868,062 connections. The 101,079,336 raw synapses are multiplied by 0.581, the release's FlyWire-equivalence factor, giving 58,727,094 effective synapses.

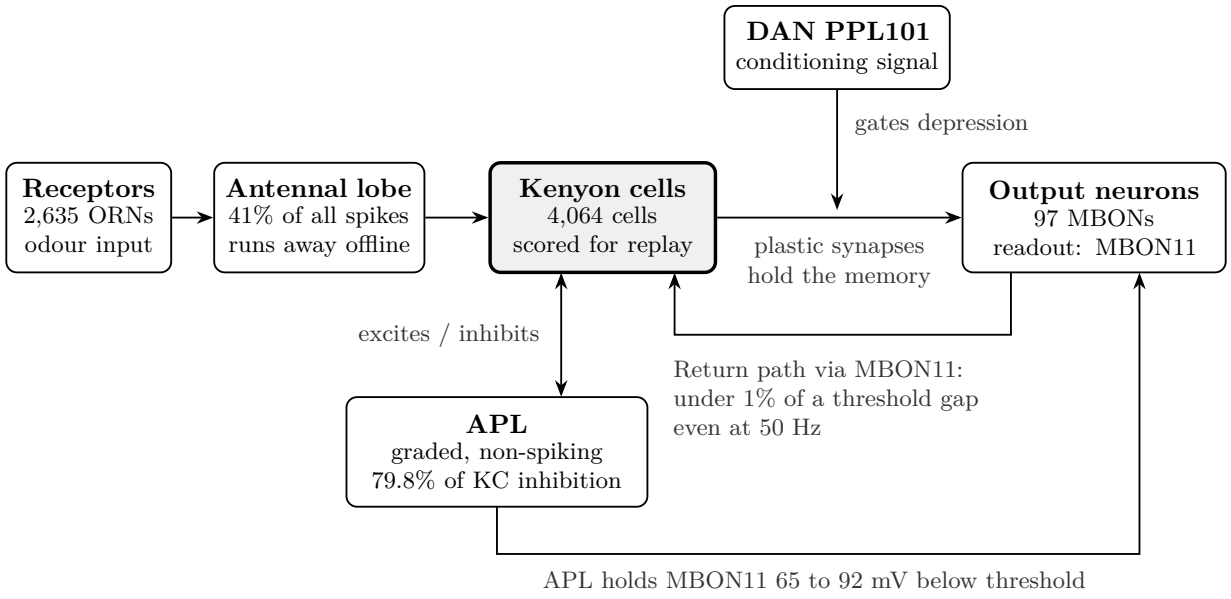


Figure 1: The memory sits one synapse downstream of the cells scored for replay. Learning depresses KC→MBON synapses, gated by the conditioning dopaminergic neuron PPL101. The engram can reach the Kenyon cells only through MBON11's return path, which is too weak to move them, and the corrected APL holds MBON11 far below threshold. Numbers are from the model at its operating point.

Each presynaptic neuron's sign comes from its predicted consensus transmitter. GABA, glutamate and histamine are inhibitory; acetylcholine, dopamine, octopamine and serotonin are excitatory. The 952 neurons with no prediction default to excitatory, a choice with no source behind it. The result is 13,502,767 excitatory and 8,365,295 inhibitory connections. FlyWire v630 and v783 appear only in the engine check and the cross-dataset control.

2.2 Neuron model

Every neuron is a single-compartment LIF unit with the published equations [28]:

$$\frac{dv}{dt} = \frac{v_{\text{rest}} - v + g}{\tau_m} + \sigma\sqrt{2/\tau_m}\,\xi(t), \qquad \frac{dg}{dt} = -\frac{g}{\tau_{\text{syn}}}. \tag{1}$$

A presynaptic spike adds the connection weight w to g after a fixed delay. A neuron spikes when v exceeds threshold, then resets and is refractory. The noise term is this project's addition (Section 2.4); the published model has none. Simulations use Brian2 in C++ standalone mode. Parameters are listed in Table 1.

2.3 Engine check and validation

The engine was first run on FlyWire v630, the published model's own data, and compared with the output file in its repository. Sugar gustatory receptor neurons (GRNs) were driven at 100 and 200 Hz and the MN9 motor neuron's rate was the benchmark. On the male CNS, the 21 published sugar neurons were mapped to labellar LB3 neurons by NBLAST and the same propagation test was run. The 0.581 scale factor was tested against 39,307 cell-type pairs present in both datasets, and fourteen target populations were cross-matched to FlyWire in both directions.

Table 1: Neuron-model parameters.

Parameter	Value	Source
Resting and reset potential	-52 mV	Kakaria and de Bivort [17], via Shiu et al. [28]
Threshold	-45 mV (7 mV gap)	Kakaria and de Bivort [17], via Shiu et al. [28]
Membrane time constant	20 ms	Product of R and C from Kakaria and de Bivort [17]; not printed in Shiu et al. [28]
Synaptic time constant	5 ms	Jürgensen et al. [16], via Shiu et al. [28]; a model parameter, not a measurement
Refractory period	2.2 ms	Lazar et al. [20]; Kakaria and de Bivort [17], via Shiu et al. [28]
Synaptic delay	1.8 ms	Paul et al. [24], via Shiu et al. [28]
Weight per synapse	$0.275 \times 0.581 = 0.160$ mV	Shiu et al. [28], its one fitted parameter
Time step	0.1 ms	Brian2 default, as in Shiu et al. [28]

2.4 Background noise and operating point

Gaussian membrane noise was added with standard deviation σ and swept over 13 values from 0.5 to 4.0 mV, with up to 10 seeds each. Each run was classified as silent, critical, saturated or bistable using avalanche statistics [2] and a multistep-regression branching ratio [31]. Spikes were binned in integer simulation steps, because binning in seconds produced spurious empty bins in floating point. The operating point is the seed-independent σ whose Kenyon-cell rate is closest, on a logarithmic scale, to the measured spontaneous rate of about 0.1 Hz [29].

2.5 Two corrections to the mushroom body

Both corrections substitute a direct measurement on an identified cell for a prediction and apply the model's existing rules to it. Neither adds a numeric parameter.

APL is non-spiking. APL releases transmitter in proportion to depolarisation rather than in spikes [1]. It is given no threshold, and its output conductance onto each target is

$$g_{\mathrm{APL}} = w \, \frac{\tau_{\mathrm{syn}}}{t_{\mathrm{ref}}} \, \mathrm{clip}\!\left(\frac{v - v_{\mathrm{rest}}}{v_{\mathrm{th}} - v_{\mathrm{rest}}},\, 0,\, 1 \right), \tag{2}$$

so a cell held at threshold delivers exactly what a spiking synapse delivers at its refractory-limited maximum rate.

DPM is inhibitory. The connectome predicts dopamine for DPM, which the sign rule makes excitatory onto 3,989 of 4,064 Kenyon cells. DPM is Gad1-positive and drives chloride influx in mushroom-body neurons [11]; it is also reported as serotonergic [21]. It is assigned GABA.

2.6 Plasticity and memory encoding

The only plastic synapses are the 61,210 KC→MBON connections (463,640 synapses) between 4,064 Kenyon cells and 97 MBONs. Depression is gated by 340 dopaminergic neurons (DANs) through 1,408 DAN→MBON connections of at least 5 synapses. The rule is two-factor, anti-Hebbian depression with no postsynaptic term, as Hige et al. [12] found. Each synapse keeps an eligibility trace e that steps by 1 on every presynaptic Kenyon-cell spike, and each spike of a gating

DAN depresses the weight in proportion to it:

$$\frac{de}{dt} = -\frac{e}{\tau_e}, \qquad w \leftarrow w - \eta_{\mathrm{LTD}}\, e\, w_0 \ \text{ on each DAN spike}, \qquad 0 \le w \le w_0. \tag{3}$$

The eligibility time constant τ_e is 5 s [15]. The learning rate $\eta_{\mathrm{LTD}} = 0.0005$ is the one free parameter, fitted to the single-pairing EPSC endpoint of Hige et al. [12] in the same cell (Section 3.4). Potentiation is off, since Hige et al. [12] found none in this compartment.

Odour A drove four ORN glomeruli (DM1, DM4, VA2, DM2) at 150 Hz for 1 s; control odour B drove four others (DL5, DM3, VM7d, DC2). Odour A was paired five times with the conditioning DAN, PPL101, at 20 Hz for 2 s from 0.2 s after odour onset; the 20 Hz rate has no measurement behind it. Odour B was presented alone 3 s after each pairing, and the next pairing began 3 s after that. Membrane state was reset before every presentation because the network does not return to baseline on its own. An odour's ensemble is every Kenyon cell that fires at least one spike to it. Twenty seeds were run.

2.7 Sleep state

Sleep is induced by clamping the 32 dFB neurons matched to the R23E10 line [14] at 17 Hz. No dFB firing rate has been measured: Donlea et al. [7] and Pimentel et al. [25] report an on/off switch. The 17 Hz value is the up-state rate of the connected ExR1 helicon cells [8] and is an approximation. Wake runs leave dFB unclamped.

Each offline period runs 40 s at the operating noise, with no odour, after a 2 s warm-up. Membrane potentials and conductances are reset at onset so that conditioning activity cannot persist into the period and pass for replay. Weights are frozen during sleep. A third arm repeats each sleep run with unlearned weights at the same seed.

2.8 Replay statistics and pre-registered criteria

Spikes are binned at 100 ms [19]. Reactivation in each bin is the Pearson correlation between the Kenyon-cell spike-count vector and the odour-A template. Because that correlation depends on template size, the primary statistic is a z-score against 200 size-matched random Kenyon-cell ensembles drawn from the same run.

Each comparison is a paired effect across the 20 seeds, reported as the mean difference with a bootstrap 95% CI (10,000 resamples), Hedges' g, a Wilcoxon signed-rank p and a sign-flip permutation p. A positive replay claim requires *all four* of the following, each with a CI excluding zero in the predicted direction:

1. odour-A ensemble above the odour-B ensemble, in sleep;
2. sleep above wake, for the odour-A ensemble;
3. the real connectome above a degree-preserving shuffled connectome;
4. the odour-A ensemble above size-matched random ensembles.

A positive that fails comparison 3 is an artifact of wiring, not replay. Two comparisons are reported as additional and cannot change the verdict: learned against unlearned weights, on the z-score and on the share of offline spikes. Arms are flagged as unmatched when their ensemble sizes or Kenyon-cell rates differ by more than 1.5 times.

A bin counts as a reactivation event when its template correlation exceeds a threshold set from the size-matched random ensembles of the same run, the 95th percentile of their per-bin 95th

Table 2: Positive control and labelled deviations from the published model.

Stage	Mechanism	Constants	Status of the constants
11	Recurrent synapses within the trained ensemble multiplied by a factor (injected positive control)	Factor 1 to 300; σ 1.4 to 1.55 mV	Chosen by the experimenter; not a result
12	Spike-triggered adaptation in every neuron: a hyperpolarising term that steps by b per spike and decays with τ	$\tau \in \{200, 600, 2000\}$ ms $\times$ $b \in \{0.05, 0.15, 0.5, 1.5\}$ mV	Unsourced, so scanned
13	Short-term depression at excitatory synapses	f 0.78 / τ 893 ms; f 0.75 / τ 1,566 ms; f 0.72 / τ 2,400 ms	Measured [18, 22, 23]; scope scanned
14	SK-type afterhyperpolarisation on APL	$\tau = 491.1 \pm 72.2$ ms	τ measured [5]; gain unmeasured, scanned

percentiles. Sequence order is scored within events that contain at least four ensemble members with distinct template ranks, by Spearman correlation against a cell-identity shuffle [10].

What pre-registered means here. There was no external registry. The four comparisons, their directions and the all-four rule were written into the analysis script and its configuration and committed to the project repository (commit `eace9d4`, 11 September 2026) before any replay run existed, and they have not been changed since. Two things did change after that commit. The statistic was switched from the raw template correlation to the size-matched z-score later the same evening (commit `6fd2a6a`), because ensemble sizes differ between odours and between the real and shuffled networks. That switch also predates the first run of the replay stage, so it was made without sight of a replay score, and three of the comparisons are also reported on the raw correlation in the repository's `results.md`. The model, by contrast, changed after results were in: a first complete pass, on the model before the two mushroom-body corrections, was also negative, and the corrections and the present operating-point rule were introduced after it. The results below are from the corrected model.

2.9 Positive control and labelled deviations

Each mechanism in Table 2 adds something the published model does not have. Results from them are never presented as properties of that model. An episode is a repeated crossing of an on threshold by the population rate; a run whose population fires below 0.02 Hz per cell, one spike per cell per 40 s run, is marked sparse and cannot count as episodic.

3 Results

3.1 The engine reproduces the published model, and the network is bistable

On FlyWire v630 the reimplementation matches the published output to within 0.7% in MN9 rate, with per-neuron rate correlations of 0.9997 and 0.9999 (Table 3). The output file shipped with the published repository was generated at 200 Hz, not at the 150 Hz default in its current code, so the comparison was made at the two rates for which shipped outputs exist.

Table 3: Engine check against the published model on its own data (FlyWire v630). Active neurons are counted over all 30 trials.

	MN9 rate (Hz)		Active neurons		Per-neuron
Sugar-GRN drive	This work	Published	This work	Published	rate r
100 Hz	67.47	67.03	428	404	0.9997
200 Hz	92.80	93.27	447	448	0.9999

Table 4: Kenyon-cell odour responses with APL as published and as measured. Measured values from Turner et al. [29].

APL model	KCs responding	Rate during odour	Rate after odour	APL output
Spiking, as published	52.6%	23.05 Hz	18.06 Hz	305.7 Hz spikes
Graded, corrected	4.5%	0.79 Hz	0.13 Hz	graded, no spikes
Measured in flies	$6 \pm 5\%$	–	about 0.1 Hz	non-spiking

On the male CNS, driving the male counterparts of the sugar neurons at 200 Hz made MN9 fire at 80.4 Hz in every trial, with 431 neurons active per trial; FlyWire gives 92.8 Hz and 402 neurons per trial at the same drive. Seventeen of the 21 published sugar neurons had one-to-one matches; the other four were reported rather than forced. Thirteen of 14 target populations matched cleanly in both directions, and MN9 was ambiguous.

The 0.581 scale factor does real work: without it the same drive activates 11,064 neurons per trial instead of a few hundred. The pre-registered check on it, a count-weighted median of per-pair synapse ratios, gave 0.795, outside the declared window of 0.45 to 0.75. A total-input ratio over the same pairs gave 0.547, close to 0.581. The male CNS detects about 45% more connected pairs, mostly weak ones, so a per-pair average understates the difference in total synaptic input, which is what the model integrates. Both statistics are reported and nothing was retuned.

The published model has no noise and a baseline rate of exactly 0 Hz. Gaussian membrane noise reproduced its target standard deviation to within 0.5% with synapses off. Sweeping it found no critical regime. Instead the network is bistable: at 1.45 and 1.5 mV, seeds of the same simulation either stayed quiescent or ignited into a high-rate state, with nothing in between. Ignition probability rose from 0% at 1.4 mV to 33% at 1.45 mV, 67% at 1.5 mV and 100% at 1.55 mV and above. The operating point for all later stages is $\sigma = 1.55$ mV, where Kenyon cells fire at 0.380 Hz and the whole brain at 1.637 Hz per neuron; at 1.4 mV, the next candidate, Kenyon cells fire at 0.00008 Hz. The brain's activity never pauses at this point, so avalanche statistics do not apply. That is a property of a model with no adaptation and no short-term depression, which has nothing to hold it at an intermediate level.

3.2 Correcting APL restores a sparse odour code

Modelling APL as non-spiking moves the odour code onto the measured values, with nothing else changed (Table 4): the fraction of Kenyon cells responding falls 12-fold, their rate during the odour 29-fold and their rate after it more than 100-fold. Connectome, parameters, stimulus and seeds are identical between the two model rows; the only difference is whether APL may fire.

A per-spike account does not explain the difference: six Kenyon-cell spikes carry APL to threshold, and one APL spike removes about a sixth of a threshold gap (1.07 mV) from each Kenyon cell. Nor does the total amount of inhibition. APL collects from 4,063 Kenyon cells, so at any Kenyon-cell rate above a fraction of a hertz it is driven far past threshold in either model. Given

a threshold, it fires at 305.7 Hz, 67% of its refractory ceiling of 455 Hz. Without one, its release sits at the ceiling of Equation 2, which is 1.5 times as much inhibition in total, while the fraction of Kenyon cells responding changes 12-fold. What the correction changes is that the inhibition is continuous instead of arriving in pulses that Kenyon cells can fire between. APL carries 79.8% of all inhibition onto Kenyon cells, so this gate, saturated in both models, is the only feedback that keeps the code sparse.

The DPM correction is inert in practice. The corrected APL delivers a steady −290 mV of drive to each of the two DPM cells, through 1,595 synapses in all, 41 times DPM's threshold gap, so DPM never fires under either transmitter; both runs give 4.45% of Kenyon cells at 0.778 Hz. The corrected APL silences the readout output neuron, MBON11, in the same way. Both cells respond to odour in real flies, so on this point the corrected model does not reproduce them either.

Two odours leave different but overlapping ensembles. At the published gain, odour A recruits 4.3% of Kenyon cells and odour B 1.5%, with a Jaccard overlap of 0.278 against 0.011 expected by chance. Scaling all synaptic weights down, a labelled deviation scanned from 0.1 to 1.0 times, did not help: at 0.6 times and below no Kenyon cell responded at all, and above 0.1 times the rest of the network never returned to baseline after the odour.

3.3 The offline state is the antennal lobe running away

Odour input does not drive a response and stop; it triggers a transition the network never leaves. One receptor neuron driven at 50 Hz is enough to ignite the whole network, and the population rate after the odour equals the rate during it. This runaway belongs to the model, not to the male CNS. Pinned to the published model, olfactory input ignited every network tested in 100% of conditions, including FlyWire v630 and v783. The gustatory pathway, with the same code and parameters, never ignited the male CNS and ignited FlyWire only when all of its 112 to 122 labellar neurons were driven at 150 Hz. The 21-neuron stimulus of the published benchmark stays below that, so the benchmark could not reveal the instability.

At the operating point, 41% of all spikes come from the 3,779 antennal-lobe neurons, 2.6% of the brain. Its local interneurons fire at 122 Hz with 96% active, and its projection neurons at 75 Hz with 93% active. Under the sign rule, 73% of local-interneuron synapses onto each other are excitatory, so the population that damps this pathway in the fly amplifies it here. Kenyon cells contribute only 0.7% of the brain's spikes.

The Kenyon cells are driven from upstream, not by each other. Of their 1,153,845 recurrent synapses, 90.0% sit on axons in the lobes and pedunculus and only 2.9% in the dendritic calyx. Removing the part of it in the lobes that lands on α/β and γ Kenyon cells, 763,745 synapses and 37% of all Kenyon-cell input, changed Kenyon-cell rates by −12.5% and the whole brain by −0.2%. The network stayed bistable.

3.4 A memory is written at the synapse, but the readout cannot show it

The plasticity rule behaves as specified. On an isolated three-neuron circuit, pairing Kenyon-cell activity with dopamine depressed the synapse to 0.000 of its starting weight, while dopamine alone or Kenyon-cell activity alone left it unchanged. In the full network the spike readout cannot be calibrated: MBON11 does not fire to the odour before or after pairing, so the graded reduction of about 80% that Hige et al. [12] measured in its spike response is not reproducible here. The learning rate was fitted instead to the other endpoint they measured after the same single pairing, the odour-evoked EPSC, taken here as the summed weight of the plastic synapses from the Kenyon cells the odour drove. The lowest rate in the grid, 0.0005, leaves 0.057 of it against a target of 0.10.

Table 5: Replay comparisons at published parameters (20 seeds, 100 ms bins). Differences are z-scores against size-matched random ensembles, except comparison 4, which is in correlation units. p is the sign-flip permutation p.

Comparison	Counted	Difference	95% CI	g	p	Predicted effect
1. Odour A vs odour B ensemble, sleep	yes	-9.097	$[-10.647, -7.593]$	-2.43	< 0.001	no
2. Sleep vs wake, odour A	yes	-0.032	$[-0.056, -0.007]$	-0.53	0.025	no
3. Real vs shuffled connectome	yes	-17.637	$[-19.205, -16.152]$	-4.74	< 0.001	no
4. Odour A vs random ensembles	yes	$+0.563$	$[+0.560, +0.566]$	$+77.08$	< 0.001	yes
Learned vs unlearned weights	additional	$+0.016$	$[-0.011, +0.044]$	$+0.24$	0.29	no
Learned vs unlearned, share of offline spikes	additional	$+0.038$	$[+0.013, +0.064]$	$+0.60$	0.011	yes

The pre-registered spike-level test of learning has no signal. MBON11 fires at 0.00 Hz before and after conditioning, because the corrected APL holds it below threshold. Learning was therefore verified at the synapse, in the same cell after the same pairing. Synapses from Kenyon cells that responded to neither odour kept 0.996 of their weight, those driven only by odour A kept 0.019, and those driven by both kept 0.001, so the depression is confined to cells an odour drove. It separates the two odours only weakly. The few synapses driven only by the unpaired odour B, about 11 per seed against 163 for A, kept 0.061 (paired over 20 seeds; Hedges' $g = -0.65$; 95% CI of the difference -0.069 to -0.016; $p = 0.0071$), so the control odour was depressed almost as far as the trained one. The likely route is eligibility carried over, since the 5 s trace outlasts the 3 s between each unpaired presentation of B and the next pairing. An engram on the odour-A ensemble, which is what the replay test needs, is present either way.

The two ensembles are not disjoint. Odour A recruits 178 Kenyon cells on average and odour B 61, and 53 of B's cells (86%) also respond to A. The A-against-B comparison is therefore largely a set against a subset of itself, and the comparison against size-matched random ensembles carries more weight.

3.5 The replay test fails at published parameters

The pre-registered test finds no evidence of memory replay, and no basis for calling what it measured an artifact either (Table 5). Only one of the four required comparisons shows the predicted effect.

Two checks came first. The sleep clamp works on its target: dFB neurons fire at 17.31 Hz in sleep and 2.64 Hz in wake. But it barely reaches anything else, changing the population rate by 0.2% (1.8676 against 1.8632 Hz) and the Kenyon-cell rate by 0.2% (0.4325 against 0.4318 Hz). And the memory does reach the Kenyon cells: offline spike trains with learned and unlearned weights differ in all 20 seeds, with active-cell overlaps of 0.84 to 0.93. Ten of the 17 MBONs that the conditioning DAN gates fire offline and project back onto Kenyon cells, through 1,732 synapses.

Three of the failures say little about replay. Comparison 1 sets odour A against a smaller ensemble that is mostly a subset of it and whose own synapses were depressed almost as far (Section 3.4), and the two differ 2.9 times in size. Comparison 2 asks for a sleep–wake difference the model does not have, since the clamp moves the rest of the brain by 0.2%. Comparison 3 sets arms that differ 3.8 times in ensemble size (178 against 670 cells) and 10.7 times in Kenyon-cell rate; z-scoring corrects size but not state, so its result is not evidence in either direction.

The decisive contrast is the additional one: the identical sleep run with unlearned weights.

On the primary measure the memory contributed nothing (+0.016, CI −0.011 to +0.044). The memory changes offline activity, but the change is not an increase in how much the trained ensemble reactivates. The spike-share version of the same contrast is positive here, but it was not pre-registered, and Section 3.7 shows that it does not survive in the adapted model.

The measure itself is continuous rather than episodic: 99% of 100 ms bins count as reactivation events, so the events are not discrete episodes. Re-analysis at 10 ms bins did not change the verdict. Within the 3,999 events that qualified for sequence scoring, firing order correlated weakly with the order of the odour response, a mean absolute rank correlation of 0.122 against 0.087 for a cell-identity shuffle.

3.6 Why there is no replay

Four measurements explain the negative: the memory is present offline but tonic; the path that could carry it back to the Kenyon cells is too weak; the readout that holds it is silenced; and the model cannot hold an episode even when one is supplied.

The memory is legible offline, and tonic. Drive through the learned synapses during sleep was 18,664 mV/s in the trained network against 27,659 mV/s with unlearned weights, a change of $-32.5 \pm 1.3\%$ (paired $t = -111.7$, $p = 3.1 \times 10^{-28}$, 20 seeds). The weights themselves moved by only −1.23% of total plastic weight, so offline activity amplifies the engram about 26 times. Almost all of the change (−32.8%) is learned weights acting on unchanged activity; changed activity on unlearned weights adds +0.4%. Across 39 one-second bins the trained-to-unlearned ratio has a coefficient of variation of 0.23. The memory is expressed continuously, so it is an offline memory trace, not replay.

The return path is wired but far too weak. MBON11 is the only engram-carrying output neuron whose Kenyon-cell targets are enriched for the trained ensemble, by 1.86 times; every other carrier that fires offline contacts the ensemble less than chance. One MBON11 spike delivers 0.228 mV to a target, so even at 50 Hz it supplies 0.057 mV, 0.82% of the 7 mV threshold gap. This ceiling follows from anatomy and the model's constants, not from a run. MBON11 fired in only 1 of 20 seeds. There the ensemble sped up by 7.6%, but untouched Kenyon cells sped up by 7.9% and the whole population by 7.7%: a global shift, not disinhibition. In the 19 silent seeds the ensemble moved +0.02% ($p = 0.68$).

The readout is silenced by APL. Offline, the two MBON11 cells sit 65 and 92 mV below threshold, 9 and 13 threshold gaps, and their 95th-percentile potentials are still 62 and 88 mV short (membrane potential recorded for 8 s in one seed). APL alone accounts for 100 and 103% of those deficits, through 180 and 261 synapses. APL has about 230,000 input synapses and is modelled as one isopotential compartment, which Amin et al. [1] showed it is not. The published data contain a parameter-free partition: APL's synapses divide across 36 named neuropils, the largest holding 15.2% of its input. Splitting APL along them would cut the drive reaching any one compartment at least 6.6 times. That experiment is made possible here but was not run.

The model cannot hold an episode. No detector in this project had been shown to find replay when replay is present, so none could certify its absence. The positive control supplied the missing mechanism directly: recurrent synapses within the trained ensemble were multiplied by a factor from 1 to 300, over 18 runs at noise levels of 1.4, 1.47, 1.5 and 1.55 mV. Not one run produced

Table 6: Replay comparisons with adaptation ($\tau = 200$ ms, $b = 0.5$ mV; 20 seeds). Units and p as in Table 5.

Comparison	Counted	Difference	95% CI	g	p	Predicted effect
1. Odour A vs odour B ensemble, sleep	yes	-10.627	$[-12.751, -8.735]$	-2.16	< 0.001	no
2. Sleep vs wake, odour A	yes	$+0.056$	$[+0.003, +0.112]$	$+0.42$	0.066	yes, by CI
3. Real vs shuffled connectome	yes	-26.503	$[-28.015, -25.168]$	-7.58	< 0.001	no
4. Odour A vs random ensembles	yes	$+0.275$	$[+0.273, +0.276]$	$+86.27$	< 0.001	yes
Learned vs unlearned weights	additional	$+0.040$	$[-0.010, +0.090]$	$+0.33$	0.144	no
Learned vs unlearned, share of offline spikes	additional	$+0.039$	$[-0.016, +0.100]$	$+0.27$	0.234	no

a repeated episode. The ensemble was silent, on continuously, or switched on once and never off, and the noise rather than the injection decided which. Nothing in the published model is slower than its 20 ms membrane time constant, so nothing can end an up state. This is a property of the model class, not evidence that the fly does not replay.

3.7 A slow variable produces episodes, but still no replay

Everything in this section is a labelled deviation from the published model.

Adaptation gives the network episodes. With spike-triggered adaptation in every neuron, 1 of the 12 scanned settings gave the whole brain repeated up and down states, and 4 gave them to the trained ensemble. These were the first repeated episodes anywhere in the project, which confirms the positive control's diagnosis from the other side: the missing ingredient was a slow variable. Neither constant is measured, so the range is the result, not any single setting.

The pre-registered test on a model that can hold an episode. The full test was repeated at $\tau = 200$ ms and $b = 0.5$ mV: 80 whole-brain 40 s runs, 20 seeds in each of sleep, unlearned weights, wake and shuffled connectome. The mushroom body became more realistic. Kenyon-cell activity fell from 0.43 to 0.049 Hz and about 4% of Kenyon cells were active, close to the 5 to 6% measured in flies. The trained ensemble fired at 1.08 Hz against 0.049 Hz for Kenyon cells as a whole. The memory still contributed nothing (Table 6).

For the first time, the learned-against-unlearned comparison is matched on both ensemble size and Kenyon-cell rate, and it finds no effect. The ensemble still beats size-matched random ensembles by a wide margin, but it does so with the memory erased too, so that margin is the ensemble's own excitability rather than a memory. Comparison 2, sleep against wake, now meets the criterion by its confidence interval, with a permutation p of 0.066, so two of the four required comparisons show the predicted effect; all four are required, and the verdict is unchanged. The shuffled arms still differ 3.8 times in size and 26.7 times in Kenyon-cell rate, so comparison 3 remains uninformative. And 98.9% of time bins still count as reactivation events: episodes in the ensemble's rate are not episodes in the replay measure.

Episodes are real where they occur, and rare. A census of all 20 seeds found the trained ensemble on continuously, as a single 40 s episode, in 13 (Figure 2). The other 7 burst: six at 1.7 to 2.6 Hz, with 69 to 103 episodes per run, and one, seed 1, at 0.80 Hz with 32. Only seed 1 falls

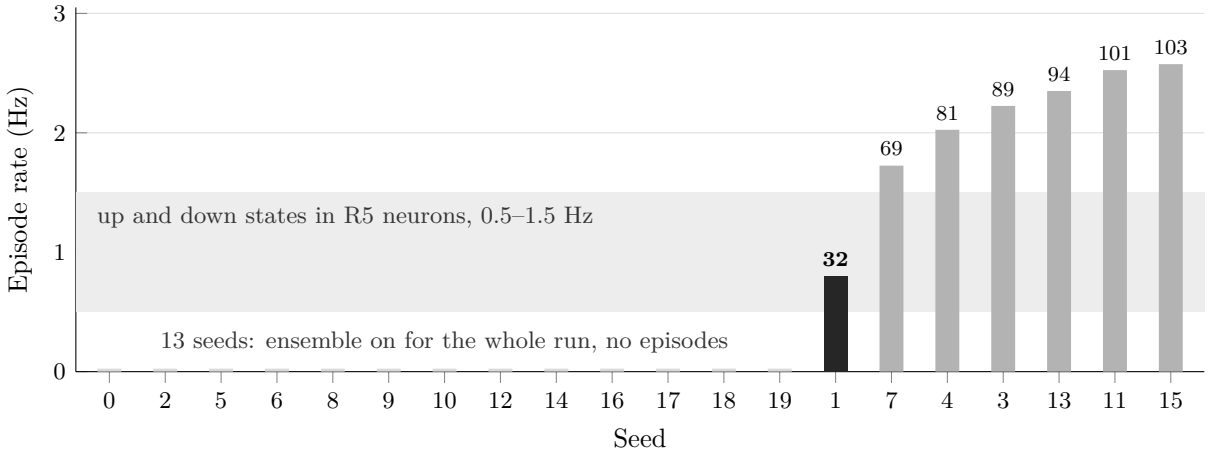


Figure 2: Only 1 of 20 seeds bursts inside the band measured in flies. Bars show the trained ensemble's episode rate in each 40 s sleep run with adaptation ($\tau = 200$ ms, $b = 0.5$ mV); numbers above bars are episode counts. Seeds on the left had one continuous episode for the whole run. Band from Raccuglia et al. [26].

inside the 0.5 to 1.5 Hz band of up and down states measured in the fly's R5 neurons [26]. An earlier look at three seeds had suggested episodic behaviour was typical; the census shows it is not.

Measured short-term depression removes the runaway, and the mushroom-body activity with it. The one slow mechanism whose magnitude is measured is depression at antennal-lobe synapses, fit three times, onto two postsynaptic cell types. What is not measured is which synapses to apply it to, so scope was an arm and all nine runs (three fits $\times$ three scopes) are reported. Where it was measured, at receptor-neuron synapses, it does nothing: local interneurons still fire at 126.6 Hz, because offline there is no odour and those synapses carry almost no traffic. Extrapolated to the antennal-lobe loop, it works as predicted before the runs: the local interneurons fall to between 0.2 and 3.8 Hz, and Kenyon-cell activity falls with them, from 0.43 to 0.003 Hz, a factor of about 140. The offline Kenyon-cell activity measured in every earlier stage was therefore the antennal-lobe runaway driving it; with the runaway gone there is nothing left in the mushroom body to reactivate. Under the sparsity floor, no run produced an episode in the brain or the ensemble.

A slow afterhyperpolarisation on APL, first run. Chen et al. [5] recorded an SK-channel afterhyperpolarisation on APL with a decay constant of 491.1 ± 72.2 ms, enhanced by sleep deprivation and reduced by recovery sleep; the same recordings independently confirm that APL is non-spiking. On APL this term would release the mushroom body's gain control with a lag of about half a second, shorter than the 0.67 to 2 s period implied by the measured up-state band. Its amplitude is not measured. So far only a 3 s smoke run exists, at the measured τ, one gain and one seed, and it produced no episode. The planned scan over gain and over the edges of τ's error bar has not been run, so this mechanism is untested rather than rejected.

4 Discussion

The negative is architectural. The memory in this model is a depression of KC→MBON synapses, one stage downstream of the Kenyon cells whose reactivation is scored (Figure 1). Learn-

ing can change what the mushroom body sends, and it does: offline drive through the learned synapses falls by a third. It cannot change which Kenyon cells switch on unless the change travels back. The only return path that prefers the trained ensemble delivers under 1% of a threshold gap even at 50 Hz, and its source is silenced by APL in 19 of 20 seeds. Adding episodes did not alter this, because episodes do not move the engram. In this model class a positive Kenyon-cell replay result would have to be explained by something other than the memory.

A model can pass its benchmark and still be wrong elsewhere. The published model reproduces its gustatory benchmark closely, and this reimplementation matches it. Yet in olfaction it codes odours with more than half of all Kenyon cells, because APL is given a threshold it does not have: inhibition that should be continuous arrives in pulses that Kenyon cells can fire between. The benchmark could not reveal this, because the 21-neuron gustatory stimulus never ignites the network. Whole-brain models are validated pathway by pathway, and a non-spiking cell modelled as spiking can be invisible in every pathway but its own.

Measured values beat fitted ones here. Both corrections, and the depression constants, came from recordings on identified cells and added no parameters. The APL correction alone moved the code onto the measured sparseness. By contrast, the natural global knob, scaling every weight down, only silenced the Kenyon cells before it stopped the runaway. The pattern suggests that the errors in connectome models are often in cell-type physiology and transmitter assignment rather than in global constants.

What a replay-capable model would need. The measurements point to four requirements. First, a slow variable that ends an up state; the published model has nothing slower than 20 ms. Second, a background that pauses rather than a runaway, which here means resolving why antennal-lobe local interneurons are 73% excitatory onto each other under the sign rule. Third, plasticity at or upstream of the Kenyon cells, or a return path strong enough to carry the engram back. Fourth, an APL that does not silence the readout, for which the connectome already provides a parameter-free partition into 36 neuropil compartments. A sleep manipulation that changes more than 0.2% of brain activity would also be needed for the sleep-against-wake comparison to mean anything.

The tonic trace is the model's real offline memory signal. The engram is expressed continuously, amplified about 26 times relative to the weight change, because offline activity falls disproportionately on the synapses that conditioning depressed. This is not replay by any of the criteria used here. It is, however, a concrete form in which a downstream engram can be present during sleep without any ensemble reactivating. Whether flies show an analogous offline shift in MBON input is an open question that this model raises but cannot settle.

Method. Three choices decided the interpretation. The unlearned-weights control, not one of the four pre-registered comparisons, was the only test that asked whether the memory contributed anything. Arm matching exposed that two of the four comparisons set networks in different states against each other. And the positive control showed that, without a slow variable, no detector could have found replay, so the first negative could not be read as evidence of absence. A census across all 20 seeds then corrected a three-seed impression that episodes were typical.

5 Limitations

The model inherits the simplifications of the published one, and this project adds a few of its own.

- **Point neurons with uniform biophysics.** Every cell is one compartment with the same membrane time constant, threshold and refractory period, whatever its type or size. There is no dendritic processing and no morphology, and conduction delay is a single constant.
- **Predicted transmitters.** Synaptic signs come from a machine-learning prediction on synapse images. Two were corrected here; others, including those of the antennal-lobe local interneurons, may be wrong. The 952 neurons with no prediction default to excitatory without a source.
- **One plasticity locus.** Only KC→MBON synapses can change, which the Discussion identifies as the main reason the test cannot come out positive.
- **Unmeasured constants.** The DAN pairing rate (20 Hz), the gating threshold of 5 DAN→MBON synapses, the dFB clamp rate (17 Hz) and the background noise have no direct measurement, and the learning rate is fitted. Each is flagged in the project's parameter table.
- **Sampling of the deviations.** The adaptation and depression scans used one seed per setting; only the re-run replay test used 20. The APL afterhyperpolarisation has a single 3 s smoke run.
- **A male brain, no body.** Much of the calibration physiology was measured in females. The ventral nerve cord is excluded, and there is no behaviour or sensory feedback.
- **No event detector and no ground truth.** The fly has no described ripple analogue to trigger on, and offline reactivation of Kenyon-cell ensembles has never been observed in a real fly, so a positive result would have had no measurement to compare against.
- **The slow-wave band is not a sleep recording.** The 0.5 to 1.5 Hz band of R5 up and down states [26] was measured in head-fixed flies and in explant brains, where its power rises over the day and after sleep deprivation; it was not recorded in flies shown to be asleep. Raccuglia et al. [27] use the same band. It serves here as the nearest measured timescale for up and down states in the fly.

6 Conclusion

A whole-brain connectome model of *Drosophila*, corrected where direct measurements contradict it, does not replay a learned Kenyon-cell ensemble during simulated sleep. The answer holds at published parameters and in a deviated model that can hold episodes and codes odours at a realistic sparseness. The reason is structural: the engram sits downstream of the cells being scored, and the path back is too weak to carry it. The memory is still present offline, as a continuous 32.5% reduction in drive through the learned synapses. Testing Kenyon-cell replay in this framework needs plasticity at or upstream of the Kenyon cells, a slow variable, and an antennal lobe that does not run away. Until then, the negative is a fact about the model, not about the fly.

Code and Data Availability

Code, configurations and result files are in the hypnagogia repository (`https://github.com/ankthba/hypnagogia`). Each stage named in this paper is one script under `scripts/`, with its out-

puts under `results/`; the base configurations are under `configs/`. The summary file `results.md` is generated from those outputs by the script `scripts/make_results.py`. The male CNS v1.0 tables and FlyWire v630 and v783 are fetched by `scripts/fetch_data.py`. The model runs on Brian2 2.10.1 in C++ standalone mode.

Acknowledgments and Use of AI Tools

This work was developed with Anthropic's Claude as a collaborator on the code, the analysis and the writing. Any errors are the author's own.

References

[1] H. Amin, A. A. Apostolopoulou, R. Suárez-Grimalt, E. Vrontou, and A. C. Lin. Localized inhibition in the *Drosophila* mushroom body. *eLife*, 9:e56954, 2020. doi: 10.7554/eLife.56954.

[2] J. M. Beggs and D. Plenz. Neuronal avalanches in neocortical circuits. *Journal of Neuroscience*, 23(35):11167–11177, 2003. doi: 10.1523/JNEUROSCI.23-35-11167.2003.

[3] S. Berg et al. Sexual dimorphism in the complete *Drosophila* male central nervous system connectome. *Cell*, 189(18):5504–5526.e15, 2026. doi: 10.1016/j.cell.2026.08.015.

[4] J. A. Berry, I. Cervantes-Sandoval, M. Chakraborty, and R. L. Davis. Sleep facilitates memory by blocking dopamine neuron-mediated forgetting. *Cell*, 161:1656–1667, 2015. doi: 10.1016/j.cell.2015.05.027.

[5] C. C. Chen, Y. C. Huang, A. Ortega, R. Suárez-Grimalt, E. Tedre, E. S. Baz, Y. Wu, A. C. Lin, and S. Liu. Sleep facilitates pattern separation through SK channel-mediated sparse coding. *Current Biology*, 36(7):1633–1643.e6, 2026. doi: 10.1016/j.cub.2026.02.028.

[6] U. Dag, Z. Lei, J. Q. Le, A. Wong, D. Bushey, and K. Keleman. Neuronal reactivation during post-learning sleep consolidates long-term memory in *Drosophila*. *eLife*, 8:e42786, 2019. doi: 10.7554/eLife.42786.

[7] J. M. Donlea, D. Pimentel, and G. Miesenböck. Neuronal machinery of sleep homeostasis in *Drosophila*. *Neuron*, 81(4):860–872, 2014. doi: 10.1016/j.neuron.2013.12.013.

[8] J. M. Donlea, D. Pimentel, C. B. Talbot, A. Kempf, J. J. Omoto, V. Hartenstein, and G. Miesenböck. Recurrent circuitry for balancing sleep need and sleep. *Neuron*, 97(2):378–389.e4, 2018. doi: 10.1016/j.neuron.2017.12.016.

[9] S. Dorkenwald et al. Neuronal wiring diagram of an adult brain. *Nature*, 634:124–138, 2024. doi: 10.1038/s41586-024-07558-y.

[10] D. J. Foster and M. A. Wilson. Reverse replay of behavioural sequences in hippocampal place cells during the awake state. *Nature*, 440:680–683, 2006. doi: 10.1038/nature04587.

[11] P. R. Haynes, B. L. Christmann, and L. C. Griffith. A single pair of neurons links sleep to memory consolidation in *Drosophila melanogaster*. *eLife*, 4:e03868, 2015. doi: 10.7554/eLife.03868.

[12] T. Hige, Y. Aso, M. N. Modi, G. M. Rubin, and G. C. Turner. Heterosynaptic plasticity underlies aversive olfactory learning in *Drosophila*. *Neuron*, 88:985–998, 2015. doi: 10.1016/j.neuron.2015.11.003.

[13] K. S. Honegger, R. A. A. Campbell, and G. C. Turner. Cellular-resolution population imaging reveals robust sparse coding in the *Drosophila* mushroom body. *Journal of Neuroscience*, 31:11772–11785, 2011. doi: 10.1523/JNEUROSCI.1099-11.2011.

[14] B. K. Hulse, H. Haberkern, R. Franconville, et al. A connectome of the *Drosophila* central complex reveals network motifs suitable for flexible navigation and context-dependent action selection. *eLife*, 10:e66039, 2021. doi: 10.7554/eLife.66039.

[15] L. Jiang and A. Litwin-Kumar. Models of heterogeneous dopamine signaling in an insect learning and memory center. *PLoS Computational Biology*, 17(8):e1009205, 2021. doi: 10.1371/journal.pcbi.1009205.

[16] A.-M. Jürgensen, A. Khalili, E. Chicca, G. Indiveri, and M. P. Nawrot. A neuromorphic model of olfactory processing and sparse coding in the *Drosophila* larva brain. *Neuromorphic Computing and Engineering*, 1:024008, 2021. doi: 10.1088/2634-4386/ac3ba6.

[17] K. S. Kakaria and B. L. de Bivort. Ring attractor dynamics emerge from a spiking model of the entire protocerebral bridge. *Frontiers in Behavioral Neuroscience*, 11:8, 2017. doi: 10.3389/fnbeh.2017.00008.

[18] H. Kazama and R. I. Wilson. Origins of correlated activity in an olfactory circuit. *Nature Neuroscience*, 12:1136–1144, 2009. doi: 10.1038/nn.2376.

[19] H. S. Kudrimoti, C. A. Barnes, and B. L. McNaughton. Reactivation of hippocampal cell assemblies: effects of behavioral state, experience, and EEG dynamics. *Journal of Neuroscience*, 19:4090–4101, 1999. doi: 10.1523/JNEUROSCI.19-10-04090.1999.

[20] A. A. Lazar, T. Liu, M. K. Turkcan, and Y. Zhou. Accelerating with FlyBrainLab the discovery of the functional logic of the *Drosophila* brain in the connectomic and synaptomic era. *eLife*, 10:e62362, 2021. doi: 10.7554/eLife.62362.

[21] P.-T. Lee, H. W. Lin, Y. H. Chang, T. F. Fu, J. Dubnau, J. Hirsh, T. Lee, and A. S. Chiang. Serotonin–mushroom body circuit modulating the formation of anesthesia-resistant memory in *Drosophila*. *Proceedings of the National Academy of Sciences USA*, 108(33):13794–13799, 2011. doi: 10.1073/pnas.1019483108.

[22] K. I. Nagel and R. I. Wilson. Mechanisms underlying population response dynamics in inhibitory interneurons of the *Drosophila* antennal lobe. *Journal of Neuroscience*, 36:4325–4338, 2016. doi: 10.1523/JNEUROSCI.3887-15.2016.

[23] K. I. Nagel, E. J. Hong, and R. I. Wilson. Synaptic and circuit mechanisms promoting broadband transmission of olfactory stimulus dynamics. *Nature Neuroscience*, 18:56–65, 2015. doi: 10.1038/nn.3895.

[24] M. M. Paul, M. Pauli, N. Ehmann, S. Hallermann, M. Sauer, R. J. Kittel, and M. Heckmann. Bruchpilot and Synaptotagmin collaborate to drive rapid glutamate release and active zone differentiation. *Frontiers in Cellular Neuroscience*, 9:29, 2015. doi: 10.3389/fncel.2015.00029.

[25] D. Pimentel, J. M. Donlea, C. B. Talbot, S. M. Song, A. J. F. Thurston, and G. Miesenböck. Operation of a homeostatic sleep switch. *Nature*, 536:333–337, 2016. doi: 10.1038/nature19055.

[26] D. Raccuglia, S. Huang, A. Ender, M. M. Heim, D. Laber, R. Suárez-Grimalt, A. Liotta, S. J. Sigrist, J. R. P. Geiger, and D. Owald. Network-specific synchronization of electrical slow-wave oscillations regulates sleep drive in *Drosophila*. *Current Biology*, 29:3611–3621.e3, 2019. doi: 10.1016/j.cub.2019.08.070.

[27] D. Raccuglia, R. Suárez-Grimalt, L. Krumm, A. Ender, C. B. Brodersen, S. R. Jagannathan, M. Freire Krück, N. P. Pampaloni, C. Rauch, Y. Winter, G. Yvon-Durocher, R. Kempter, J. R. P. Geiger, and D. Owald. Network synchrony creates neural filters promoting quiescence in *Drosophila*. *Nature*, 646:667–675, 2025. doi: 10.1038/s41586-025-09376-2.

[28] P. K. Shiu, G. R. Sterne, N. Spiller, et al. A *Drosophila* computational brain model reveals sensorimotor processing. *Nature*, 634:210–219, 2024. doi: 10.1038/s41586-024-07763-9.

[29] G. C. Turner, M. Bazhenov, and G. Laurent. Olfactory representations by *Drosophila* mushroom body neurons. *Journal of Neurophysiology*, 99:734–746, 2008. doi: 10.1152/jn.01283.2007.

[30] M. A. Wilson and B. L. McNaughton. Reactivation of hippocampal ensemble memories during sleep. *Science*, 265:676–679, 1994. doi: 10.1126/science.8036517.

[31] J. Wilting and V. Priesemann. Inferring collective dynamical states from widely unobserved systems. *Nature Communications*, 9:2325, 2018. doi: 10.1038/s41467-018-04725-4.