Emergence of time persistence in an interpretable data-driven neural network model

Sebastien Wolf[a,*], Guillaume Le Goc[b,*], Simona Cocco[a,**], Georges Debrégeas[b,**], Rémi Monasson[a,c,d,**]

[a] *Laboratory of Physics of the Ecole Normale Supérieure, CNRS UMR 8023 & PSL Research, Sorbonne Université, Université de Paris, Paris, France*
[b] *Sorbonne Université, CNRS, Institut de Biologie Paris-Seine (IBPS), Laboratoire Jean Perrin (LJP), Paris, France*
[c] *Corresponding author*
[d] *Lead contact*

Abstract

Establishing accurate as well as interpretable models of networks activity is an open challenge in systems neuroscience. Here we infer an energy-based model of the ARTR, a circuit that controls zebrafish swimming statistics, using functional recordings of the spontaneous activity of hundreds of neurons. Although our model is trained to reproduce the low-order statistics of the network activity at short time-scales, its simulated dynamics quantitatively captures the slowly alternating activity of the ARTR. It further reproduces the modulation of this persistent dynamics by the bath temperature and visual stimulation. Mathematical analysis of the model unveils a low-dimensional landscape-based representation of the ARTR activity, where the slow network dynamics reflects Arrhenius-like barriers crossings between metastable states. Our work thus shows how data-driven models built from large neural populations recordings can be reduced to low-dimensional functional models in order to reveal the fundamental mechanisms controlling the collective neuronal dynamics.

Keywords: Calcium Imaging, Neuronal Network, Zebrafish, Neural Persistence, Data-Driven Ising Model, Mean Field

1 Introduction

Understanding how collective computations emerge at the neuronal population level and how they are supported by appropriate circuitry is a central goal of neuroscience. In this context, modeling efforts have long been based on top-down approaches, in which mathematical models are designed to capture the key mechanisms underlying function. While being very fruitful from a conceptual point of view, these models show a limited capability to accurately reproduce activity measurements, which

*These authors contributed equally.
**Senior authors with equal contributions.
Email addresses: sebastien.wolf@ens.fr (Sebastien Wolf), remi.monasson@phys.ens.fr (Rémi Monasson)

hinders their validation. Recently, progress in large-scale recording and simulation techniques have led to the development of bottom-up approaches, in which machine-learning-based models fit recorded data with high precision, and allow for decoding or prediction of activity and behavior (Glaser et al., 2020; Pandarinath et al., 2018). Unfortunately, the blackbox nature of these powerful data-driven models make their biological interpretation, such as the identification of relevant neural assemblies and of the computational processes they support often difficult (Butts, 2019). It is therefore important to develop quantitative, yet interpretable approaches to tackle the functions carried out by large neural populations. The present work is an attempt to do so in the specific context of the anterior rhombencephalic turning region (ARTR), a circuit in the zebrafish larva that shows slow ($\sim$ 10 sec) endogenous alternations between left and right active states driving the gaze orientation and chaining of leftward/rightward swim bouts (Ahrens et al., 2013; Dunn et al., 2016; Wolf et al., 2017; Ramirez & Aksay, 2021; Leyden et al., 2021).

The ARTR was chosen as it constitutes an experimentally accessible persistent circuit whose functional role is well identified. Persistent circuits endow the brain with the capacity to store information for finite periods of time, a feature that is essential to action selection, as it allows for the accumulation of sensory data until a consequential decision is made (Wang, 2008). It is also required for motor control, e.g. to hold the eyes at a given orientation (Seung, 1996; Seung et al., 2000). Short-term memory storage engages dedicated circuits that can maintain persistent patterns of activity in response to transient inputs (Zaksas & Pasternak, 2006; Guo et al., 2017). Different patterns may represent discrete or graded stored variables, associated with physical parameters such as head or eyes direction, or abstract percepts. As isolated neurons generally display short relaxation times, neural persistence is thought to be an emergent property of recurrent circuit architectures (Zylberberg & Strowbridge, 2017). Since the 1970s, numerous mechanistic network models have been proposed that display persistent activity. They are designed such as to possess attractor states, *i.e.* stable activity patterns towards which the network spontaneously converges.

Although attractor models are conceptually appealing, assessing their relevance in biological circuits remain challenging. To this aim, recent advances in machine learning combined with large-scale methods of neural recordings may offer a promising avenue. We hereafter focus on energy-based network models, trained to replicate low-order data statistics, such as the mean activity and pairwise correlation, through the maximum entropy principle (Jaynes, 1957). In neuroscience, such models have been successfully used to explain correlation structures in many areas, including the retina (Schneidman et al., 2006; Cocco et al., 2009; Tkačik et al., 2015), the cortex (Tavoni et al., 2016, 2017; Nghiem et al., 2018), and the hippocampus (Meshulam et al., 2017; Posani et al., 2017) of vertebrates, and the nervous system of *C. elegans* (Chen et al., 2019). These models are generative, *i.e.* they can be used to produce synthetic activity on short time scales, but whether they can reproduce long-time dynamical

features of the biological networks remains an open question.

Here, we first report on spontaneous activity recordings of the ARTR network using light-sheet based functional imaging at various temperatures. These data demonstrate that the bath temperature controls the persistence time-scale of the ARTR network, and that this modulation is in quantitative agreement with the thermal dependence of the swimming statistics (Le Goc et al., 2021). We then infer energy-based models from the calcium activity recordings, and show how these data-driven models not only capture the characteristics and probabilities of occurrence of activity patterns, but also reproduces the observed thermal-dependence of the persistent timescale. We further derive a mathematically tractable version of our energy-based model, called mean-field approximation, whose resolution provides a physical interpretation of the energy landscape, the dynamical paths there in, and their changes with temperature. We finally extend the model to incorporate visual stimulation and correctly reproduce the previously reported visually-driven ARTR dynamics (Wolf et al., 2017). This work establishes the capacity of data-driven network inference to numerically emulate persistent dynamics and to unveil fundamental network features controlling such dynamics.

2 Results

2.1 Temperature enables manipulating behavioral and neuronal persistence time-scales in zebrafish larvae

In this first section, we report on functional recordings of the ARTR dynamics performed at various temperature (18 to 33°C). We characterize how the bath temperature impacts the distribution of activity patterns as well as the time-scale of activity alternation that characterizes the ARTR's endogenous dynamics. As the ARTR functional role in zebrafish navigation is well established, we compare this thermal modulation of the ARTR neural activity with the associated change in swimming statistics measured in freely swimming assays.

2.1.1 Temperature-dependence of the orientational persistence time-scale in zebrafish larvae

We first characterize from a behavioral viewpoint how the water temperature can be used to modulate the time scale of orientational persistence in an ethological manner. Zebrafish navigation is comprised of discrete swim bouts lasting for 100-200 ms interspersed by 1-2 s-long periods of inactivity. The bouts can be categorized into two main types: forward scouts that propel the animal forward, and turn bouts, which further induce heading reorientation (Figure 1A). Orientational persistence manifests itself as the preferred chaining of similarly orientated turn bouts (Chen & Engert, 2014; Dunn et al., 2016; Karpenko et al., 2020): a leftward (respectively, rightward) turn is more likely

to be observed when the previous turn was oriented to the left (respectively, to the right). This motor-memory mechanism, which lasts over 5 bouts, has important consequences for the long-term exploratory dynamics (Karpenko et al., 2020).

Zebrafish is a cold-blooded (poïkilotherm) animal, whose internal temperature is set by the water bath. In their natural habitat, zebrafish can experience (and survive) temperatures ranging from 15 to 35°C (Engeszer et al., 2007). To study how the temperature of the bath modulates the kinematic parameters controlling the fish spontaneous navigation (Le Goc et al., 2021), we video-monitored 5-7 days old zebrafish larvae as they freely swim in a rectangular bath at constant and uniform temperature in the range 18-33°C (Figure 1A). We found that higher bath temperature increases the fraction of turn bouts, induces larger reorientation angles, shortens interbout-intervals, and increases forward displacement.

In the scope of the present study, we focus on the impact of temperature on orientational persistence. To quantify the time scale of this mechanism, we hypothesized the existence of an underlying two-state continuous process governing the selection of leftward vs rightward turn bouts. We assigned a discrete value to each turn bout, -1 for a right turn, $+1$ for a left turn, while forward scouts are ignored. The orientational state signal is then defined, at each time step, by the value of the last turn bout (Figure 1B). The power spectra of the resulting binary signals, shown in Figure 1C for various temperatures, are Lorentzian as expected for a single-time telegraph process (see Methods). The corresponding fits (Methods, Eq. 5) allow us to extract, for each experiment, a frequency k_{flip} defined as the probability of switching state per unit of time. As shown in Figure 1D, this rate systematically increases with the temperature, from 0.1 to 0.6 s^{-1}. Increasing temperature thus leads to a progressive reduction of the orientational persistence time. Notice that at the highest tested temperature (33°C), the persistence time becomes comparable to the interbout interval, such that the chaining of left and right turn becomes essentially memory-less.

2.1.2 ARTR right/left alternation dynamics is thermally modulated in line with behavior

The ARTR is a bilaterally distributed neural population located in the anterior hindbrain, which has been identified as the main selection hub for turn bouts orientation during spontaneous and visually-driven exploration (Dunn et al., 2016; Wolf et al., 2017). Leftward or rightward bouts preferentially occur when the ipsilateral subcircuit is active while the contralateral subcircuit is inactive.

Based on these prior observations, we sought to examine whether the temperature dependence of the orientational persistence observed in freely swimming assays was reflected in the endogenous dynamics of the ARTR in tethered animals. We used light-sheet functional imaging to record the brain activity at single cell resolution in zebrafish larvae expressing a calcium reporter pan-neuronally

(*Tg(elavl3:GCaMP6)*). The larvae were embedded in a small cylinder of agarose gel later introduced in a water bath whose temperature was controlled in the same range (18–33°C, see Figure 1E). From these volumetric recordings, the ARTR neurons were identified using a combination of morphological and functional criteria, as detailed in Wolf et al. (2017). The spatial organization of the selected neuronal population is displayed in Figure 1F, for all recorded animals after morphological registration on a unique reference brain (145 ± 65 left neurons, 165 ± 69 right neurons, mean ± s.d. across 13 different fish, see Suppl. Table S1). For each individual neuron, an approximate spike train $s(t)$ was inferred from the fluorescence signal using Bayesian deconvolution (Tubiana et al., 2020). A typical raster plot of the ARTR is shown in Figure 1G (recorded at 26°C), together with the mean signals of the left and right subcircuits, $m_{L,R}(t) = \frac{1}{N_{L,R}} \sum_{i \in L,R} s_i(t)$.

To allow for a direct comparison between the dynamics of the ARTR and of the orientational signal extracted from behavioral assays, we defined a binarized ARTR signal monitoring which of the two regions is more active at any given time, see Eq. 6 in Methods. This signal is shown in Figure 1H for three different temperatures for the same fish. Increasing the bath temperature results in a decrease of the mean persistence time $\overline{\tau}$ in either state. In order to quantify this effect, we performed a similar analysis as for the behavioral orientational state signal. We computed the power spectra of the binarized ARTR signals (3 to 8 animals for each temperature, see Suppl. Table S1), for the five tested temperatures and used a Lorentzian fit to extract the alternation frequency $\nu = 1/\overline{\tau}$ for each animal (Figure 1I). As shown in Figure 1J, the extracted frequency increases with temperature in line with our behavioral findings. Although ν could significantly vary across specimen at a given temperature, we found that, for a given animal, increasing the temperature induced an increase in the frequency in 87.5% of our recordings (28 out of 32 pairs of recordings). The inset plot further establishes that the left/right alternation rate extracted from behavioral and neuronal recordings are consistent (slope $= 0.81$, $R = 0.99$).

2.1.3 ARTR activity maps are modulated by the temperature

We then investigated the effect of the bath temperature on the ARTR activity patterns. At each time step, we characterized the ARTR activation state by the mean activity of the left and right sub-populations, m_L and m_R. The probability map in the (m_L, m_R) plane is shown in Figure 2A, for three different temperatures. This map shows that, concurrent with the change in the circuit dynamics (Figure 2B), temperature also impacts the distribution of activities. At high temperature, the ARTR activity map is mostly confined within a L-shaped region around $(m_L = 0, m_R = 0)$ and the circuit is essentially inactive for a significant fraction of the time. Conversely, lower temperature tends to increase the duration of high ARTR activity periods, leaving shorter time intervals during which both regions are inactive. To quantify the change in the activity distribution we computed the

log-probability of the activity of either region of the ARTR at various temperatures (Figure 2C). The occupation rate of the inactive state ($m_{L,R} \sim 0$) increases with temperature, with a corresponding steeper decay of the probability distribution of the activity. Consistently, we found that the mean activities m_L and m_R decreased with temperature (Figure 2D).

Our analysis thus indicates that the bath temperature modulates both the internal dynamics and the activity distribution of the ARTR. However, we observe for both aspects a large variability between animals at a given temperature. This variability in the endogenous neuronal dynamics is not unexpected, as it parallels the intra- and inter-individual variability in the fish exploratory kinematics reported in Le Goc et al. (2021). We thus asked whether the circuit persistence time scale and activity distribution consistently varied between animals and trials for a given temperature. We plotted, for all datasets, the mean persistence time as a function of the mean activity of both populations (see Figure 2E and Methods). We found a quasi-linear dependence between both quantities ($R = 0.91$), showing the strong correlation between these two characteristics of the circuit internal dynamics.

2.2 A data-driven energy-based model reproduces the statistical and dynamical properties of ARTR neural activity

2.2.1 Inference of an Ising model from low-order statistics of neural activity of the ARTR.

We used the recordings of the ARTR spontaneous activity to infer of an energy-based network model. Our approach, going from raw fluorescence data to the model, is summarized in Figure 3A. We first reconstructed the spike trains of each ARTR neuron using a deconvolution algorithm (Tubiana et al., 2020), and obtained a raster plot of the neural activity for each recording ($n = 13$ fish, $N = 100$ to 300 cells, see Suppl. Table S1). We divided the recording window T_{rec} (of the order of 1200 s for each session) in time bins of length adjusted to the imaging framerate (from 100 to 300 ms). Each data set thus consisted of a series of snapshots $\mathbf{s^k} = (s_1^k, \ldots, s_n^k)$ of the neural population activity in the bins k, with $k = 1, \ldots, t_{rec}/Dt$; here, $s_i^k = 1$ if cell i is active or $s_i^k = 0$ if it is silent in time bin k. We extracted from those data the mean activities, $\langle s_i \rangle_{\text{data}}$, and the pairwise correlations, $\langle s_i s_j \rangle_{\text{data}}$, as the averages of, respectively, s_i^k and $s_i^k s_j^k$ over all time bins k.

We then inferred the least constrained model, according to the maximum entropy principle (Jaynes, 1957), that reproduced the mean activities $\langle s_i \rangle_{\text{data}}$ and the pairwise correlations $\langle s_i s_j \rangle_{\text{data}}$. This model, known as the Ising model in statistical mechanics (Ma, 1985) and probabilistic graphical model in statistical inference (Koller & Friedmann, 2009), describes the probability distribution over all 2^N

possible activity configurations $\mathbf{s}$,

$$P(\mathbf{s}) = \frac{1}{Z} \exp\left(\sum_i h_i s_i + \sum_{i<j} J_{ij} s_i s_j\right) , \quad (1)$$

where Z is a normalization constant. The bias h_i controls the intrinsic activity of neuron i, while the coupling parameters J_{ij} account for the effect of the other neurons j activity on neuron i (Methods).

The set of parameters $\{h_i, J_{ij}\}$ were inferred using the Adaptative Cluster Expansion and the Boltzmann machine algorithms (Cocco & Monasson, 2011; Barton & Cocco, 2013; Barton et al., 2016), to best reproduce the observed mean activities and pairwise correlations (Suppl. Fig. S1B-D). In addition, the Ising model captures higher-order statistical properties of the activity such as the probability that K cells are active in a time bin (Suppl. Fig. S1E) (Schneidman et al., 2006).

Figure 3A reveals the spatial (left/right) organization of pairwise correlations: we observe strong positive correlations between ipsilateral cells, and weak negative or positive correlations between contralateral cells, yielding an overall bimodal distribution (left/right). Consistent with these experimental observations, the coupling matrices J_{ij} in the model also show a left/right organization (ttest, $p_{value} < 10^{-5}$), with the distribution of couplings between ipsilateral cells centered around a positive value (std = 0.12, mean = 0.062 over $n = 13$ fish, 32 recordings) and the one between contralateral cells centered around a null value (std = 0.10, mean = -0.001), see Figure 3B-C and Suppl. Fig. S1I. The ipsilateral couplings J_{ij} decay, on average, exponentially with the distance between neurons i and j (Suppl. Fig. S1A and J), in agreement with findings in other neural systems (Posani et al., 2018). Spatial structure is also present in contralateral couplings (Suppl. Fig. S1K). Biases display a wide distribution ranging from -8 to 0 (std = 1.1, mean = -4.1, Figure 3D and Suppl. Fig. S1F-H).

2.2.2 Monte Carlo sampling of Ising model reproduces temperature-dependent neural persistence.

Once inferred, the Ising model can be used to generate synthetic activity configurations $\mathbf{s}$ of the network. In practice we use a Metropolis Monte Carlo (MC) algorithm to sample the probability distribution $P(\mathbf{s})$ in Eq. 1. Figure 4A shows the activity maps obtained through MC sampling of the Ising models trained on the three same datasets as in Figure 2A. The time evolution of the mean activities of the left and right sub-populations are shown in Figure 4B. For these synthetic signals, we use MC rounds, *i.e.* the number of MC steps divided by the total number of neurons (Methods) as a proxy for time. Remarkably, although the Ising model is trained to reproduce the low-order statistics of the neuronal activity within a time bin only, and is therefore blind to the circuit dynamics, the generated signals still capture the main characteristics of the ARTR dynamics, *i.e.* a slow alternation between the left and right sub-populations associated with long persistence times.

To quantitatively assess how the activity distribution and persistence time are reproduced, we first compared the activity maps extracted from the data and from the Ising model using the Kullback-Leibler (KL) divergence, a classical metrics of the dissimilarity between two probability distributions. The distribution of the KL divergence computed between experimental datasets and corresponding Ising models is shown in green in 4C. These values are found to be systematically smaller than those obtained between any combination of datasets and Ising models trained on different datasets (red distribution). This result establishes that the Ising model can accurately reproduce the ARTR activity distribution associated to each specimen and temperature.

We also computed from each synthetic signal a mean persistence time, using the same procedure as for the experimental data. We found that the persistence times extracted from the data and from the MC simulations of the inferred models are strongly correlated (Figure 4D, $R = 0.84$). In particular, the persistence time of the MC dynamics capture the temperature dependence and inter-individual variability of the neural persistence time.

2.3 Mean-field study of the inferred model unveils the energy landscape underlying the ARTR dynamics

2.3.1 Mean-field approximation to the data-driven graphical model.

While our data-driven Ising model reproduces the decrease in persistence time and the associated changes in the states occupation with temperature, why it does so remains unclear. To understand what features of the coupling and local bias parameters are responsible for these collective properties we turn to mean-field theory, a powerful and mathematically tractable approximation scheme used in statistical physics to study systems with many strongly interacting components (Ma, 1985). Mean field, as its name suggests, amounts to determine self-consistent equations for the mean activities m_L and m_R of neurons in, respectively, the left (L) and right (R) regions of the ARTR (Methods).

Within mean-field theory, each neuron i is subject to

(i) a local bias H;

(ii) a strong excitatory coupling $J > 0$ from the neurons in the ipsilateral region and a weak coupling I from the neurons in the contralateral side.

Point (i) approximates the distribution of the inferred biases h_i (Figure 3D and Suppl. Fig. S1F-H) with their average value H. Point (ii) merely expresses the presence of recurrent excitation within each region of the weak reciprocal interaction between left and right sides (Figure 3C and Suppl. Fig. S1I). As for the biases, the inferred ipsilateral and contralateral interactions J_{ij} are replaced with their mean values, respectively, J and I. In addition, we introduce an effective size K of each region to take into account the fact that mean-field theory overestimates interactions by replacing them with

their mean value. This effective number of neurons is, in practice, fitted to best match the results of the mean-field approach to the full Ising model predictions (see Methods), and is substantially smaller than the number N of recorded neurons.

The selection method used to delineate the ARTR populations may yield different number of neurons in the L and R regions (see Suppl. Table S1). This asymmetry is accounted for by allowing the parameters H, J and K defined above to take different values for the left and right sides. As a consequence, mean–field theory allows us to simplify the data-driven Ising model, whose definition requires $\frac{1}{2}(N_L+N_R)(N_L+N_R+1)$ parameters $\{h_i,\ J_{ij}\}$, into a model depending on seven parameters $(H_L, H_R, J_L, J_R, K_L, K_R, I)$ only (Figure 5A), whose values vary with the animal and the experimental conditions e.g. temperature (Suppl. Table S2).

2.3.2 Free energy and Langevin dynamics.

The main outcome of the analytical treatment of the model is the derivation of the so-called free energy $F(m_L, m_R)$ as a function of the average activities m_L and m_R of neurons in the left and right areas, see Eq. 17 in Methods. The free energy is a fundamental quantity as it controls the density of probability to observe an activation pattern (m_L, m_R) through

$$P(m_L, m_R) \propto e^{-F(m_L, m_R)} \tag{2}$$

Consequently, the lower the free energy F, the higher the probability of the corresponding state (m_L, m_R). In particular, the minima of the free energy correspond to persistent states of activity in which the network can be transiently trapped.

The free energy landscape can be used to simulate dynamical trajectories in the activity space (m_L, m_R). To do so, we consider a Langevin dynamics in which the two activities $m_L(t), m_R(t)$ evolve in time according to the stochastic differential equations,

$$\tau \frac{dm_L}{dt}(t) = -\frac{\partial F}{\partial m_L}\big(m_L(t), m_R(t)\big) + \epsilon_L(t) \quad , \qquad \tau \frac{dm_R}{dt}(t) = -\frac{\partial F}{\partial m_R}\big(m_L(t), m_R(t)\big) + \epsilon_R(t) \tag{3}$$

where τ is a microscopic time scale, and $\epsilon_L(t), \epsilon_R(t)$ are white noise 'forces', $\langle \epsilon_L(t) \rangle = \langle \epsilon_R(t) \rangle = 0$, independent and delta-correlated in time: $\langle \epsilon_L(t)\epsilon_R(t') \rangle = 0$, $\langle \epsilon_L(t)\epsilon_L(t') \rangle = \langle \epsilon_R(t)\epsilon_R(t') \rangle = 2\,\delta(t-t')$. This Langevin dynamical process ensures that all activity configurations (m_L, m_R) will be sampled in the course of time, with the expected probability as given by Eq. 2. In the following, we will refer to a high (resp. low) level of activity as m^{high} (m^{low}, respectively).

Figure 5B shows the mean-field simulated dynamics of the left and right activities, m_L and m_R, with the parameters corresponding to three Ising models at three different temperatures, see Figure 4A. These time traces are qualitatively similar to the ones obtained with the full inferred Ising model and in the data (see Figures 4B and 2B), comprising transient periods of self-sustained activity of either sub-circuit in particular at low and intermediate temperatures. At low temperature, the asymmetric states

(m^{high}, m^{low}) and (m^{low}, m^{high}) are dynamically stable for some time, see time trace 1 in Figure 5B. At high temperature, high activity in either (left or right) area can be reached only transiently, see traces 2 and 3 in Figure 5B. These results indicate that our mean-field model is able to retain essential features of the dynamical behavior of the ARTR neural activity observed in experiments.

2.3.3 Barriers in the free-energy landscape and dynamical paths between states.

We show in Figure 5C the free-energy landscape in the (m_L, m_R) plane for the same three conditions as in Figure 5B. The minimization conditions $\frac{\partial F}{\partial m_L} = \frac{\partial F}{\partial m_R} = 0$ provide two implicit equations over the activities m_L^*, m_R^* corresponding to the preferred states. For most data sets we found four local minima: the low-activity minimum $(m_L^*, m_R^*) = (m^{low}, m^{low})$, two asymmetric minima, (m^{high}, m^{low}) and (m^{low}, m^{high}), in which only one subregion is strongly active, and a state in which both regions are active, (m^{high}, m^{high}). The low-activity minimum (m^{low}, m^{low}) is the state with lowest free energy, hence with largest probability, while the high-activity state (m^{high}, m^{high}) has much higher free energy and much lower probability. The free energies of the asymmetric minima (m^{high}, m^{low}) and (m^{low}, m^{high}) lie in between, and their values strongly vary with the temperature. The difference in free energy between the asymmetrical states and the low-activity state is an increasing function of temperature (Figure 5D-E). As a result the asymmetrical states have low probabilities compared to the low-activity state at high temperature, but are often visited at low temperature. These results are in qualitative agreement with the Langevin trajectories in Figure 5B, and with the experimental description of the ARTR states reported in Figures 2B and Figure 4B.

The Langevin dynamics defines, in addition, the most likely paths (see Methods) in the activity plane joining one preferred state to another, e.g. from (m^{high}, m^{low}) to (m^{low}, m^{high}) as shown in Figure 5C. Along these optimal paths the free energy F varies with the reaction coordinate, conveniently chosen as $m_L - m_R$, and defines barriers (Figure 5D), which have to be overcome in order for the network to dynamically switchover. The theory of activated processes tells us that the average time to cross a barrier depends exponentially on its height ΔF:

$$t(\Delta F) \sim \tau \times e^{\Delta F} \ , \tag{4}$$

up to proportionality factors of the order of unity (Langer, 1969). Thus, the barrier $\Delta F\big((m^{high}, m^{low}) \rightarrow (m^{low}, m^{low})\big)$ shown in dark green in Figure 5E controls the time needed for the ARTR to escape the state in which the left region is active while the right region is mostly silent, and to reach the all-low state. The barrier $\Delta F\left((m^{low}, m^{low}) \rightarrow (m^{low}, m^{high})\right)$ shown in purple in Figure 5E is related to the rising time from the low-low activity state to the state where the right region is very active, and the left one is silent. The height and position of the barriers can be easily computed within our mean-field model, giving access to precise information about the dynamical trajectories followed by the activity.

We therefore estimated the dependence in temperature of the barriers height (Figure 5E and Suppl. Fig. S2D) and the associated persistence time (Figure 5F). While substantial variations from animal to animal were observed, we found that barriers for escaping the all-low state and switching to either L, R region increase with temperature, as found in recordings (Figure 2E) and in Ising simulated activity (Figure 4D). Conversely, low temperatures produce free-energy landscapes with deeper minima in (m^{high}, m^{low}). This yields lower rates for the switching events from (m^{low}, m^{high}) to (m^{low}, m^{low}) and back, or, in other words, longer fixation times in the states in which one of the two regions is highly active.

2.4 Ising and mean-field models with modified biases capture the ARTR visually-driven dynamics

While the analysis above focused on the spontaneous dynamics of the ARTR, our data-driven approach is also capable of explaining activity changes induced by external and time-varying inputs. In order to illustrate this capacity, we decided to re-analyze a series of experiments, reported in Wolf et al. (2017), in which we alternatively illuminated the left and right eye of the larva, for periods of 15 to 30 s, while monitoring the activity of the ARTR (Figure 6A) with a 2-photon light-sheet microscope. During and after each stimulation protocol, 855 s of spontaneous activity was recorded on $n = 6$ fish. We found that the ARTR activity could be driven by this alternating unilateral visual stimulation: the right side of the ARTR tended to activate when the right eye was stimulated and vice-versa (Figure 6B).

To analyze these datasets we first follow the approach described in Figure 3, and inferred, for each fish, the sets of biases h_i and interactions J_{ij} using the spontaneous activity recording only (Suppl. Fig. S3C-D). In a second step, we exploited recordings of the visually-driven activity to infer additional biases δh_i to the neurons, while keeping the interactions J_{ij} fixed (Figure 6C); in practice we defined two sets of additional biases, $\delta h_i^{\leftarrow}$ and $\delta h_i^{\rightarrow}$, corresponding, respectively, to leftward and rightward illuminations. The underlying intuition is that biases encode inputs due to the stimulation, while the interactions between neurons can be considered as fixed on the experimental time scale.

The inferred values of the additional biases, averaged over the entire sub-population (right or left), are shown in Figure 6D for the ipsilateral, *e.g.* $\delta h_i^{\leftarrow}$ with i in the L subregion, and contralateral stimulations, *e.g.* $\delta h_i^{\leftarrow}$ with i in the R subregion. The results show that light stimulation produces a strong increase of excitability for the ipsilateral neurons and a smaller one for contralateral neurons.

We then simulated the visual stimulation protocol by sampling the Ising model while alternating the model parameters, from $\{h_i + \delta h_i^{\rightarrow}, J_{ij}\}$ to $\{h_i + \delta h_i^{\leftarrow}, J_{ij}\}$, and back. The simulated dynamics of the model (Figure 6E) qualitatively reproduces the experimental traces of the ARTR activity (Figure 6B). In particular, the model captures the stabilizing effect of unilateral visual stimuli, which results in

a large activation of the ipsilateral population, which in turn silences the contralateral subcircuit due to the negative I coupling between both. This yields the anti-correlation between the left and right side that is clearly visible in both the experimental and simulated traces, and much stronger than for spontaneous activity (Suppl. Fig. S3A-B-E-F).

To better understand the Ising dynamics under visual stimulation we resort, as in the spontaneous activity case studied previously, to mean-field theory. For asymmetric stimulation our mean-field model includes, during the periods of stimulation, extra biases ΔH_L and ΔH_R over neurons in, respectively, the left and right areas (Figure 6F), while the couplings J and I remain unchanged. We show in Figure 6G the free-energy F as a function of m_L, m_R for an example fish. Due to the presence of the extra bias the landscape is tilted with respect to its no-stimulation counterpart (Figure 6G), entailing that the left- or right-active states are much more likely, and the barrier separating them from the low-low state is much lower (Figure 6H). As a consequence, the time necessary for reaching the high-activity state is considerably reduced with respect to the no-stimulation case, see Eq. 4. These results agree with the large probability of the high-activity states and the fast rise to reach these states in the Ising traces in Figure 6E, compare with Figure 4B. The presence of high-activity states is seen in the traces generated from the mean-field Langevin dynamics defined by Eq. 3, alternating the bias parameter of the model, from $H + \Delta H_L$ to $H + \Delta H_R$, see Figure 6I.

3 Discussion

Modelling high-dimensional data, such as extensive neural recordings, imposes a trade-off between accuracy and interpretability. Although highly sophisticated machine-learning methods may offer quantitative and detailed predictions, they might in turn prove inadequate to elucidate fundamental neurobiological mechanisms. Here we introduced a data-driven network model, whose biologically-grounded architecture and relative simplicity make it both quantitatively accurate and amenable to detailed mathematical analysis. We implemented this approach on functional recordings performed at various temperature of a key population of neurons in the zebrafish larvae brain, called ARTR, that drives the orientation of tail bouts and gaze (Dunn et al., 2016; Wolf et al., 2017; Ramirez & Aksay, 2021; Leyden et al., 2021).

First, we demonstrate that the persistent timescale of the ARTR endogenous dynamics decreases with the temperature, mirroring the thermal modulation of orientational persistence in freely-swimming behavioral assays. We then demonstrate that our energy-based model not only captures the statistics of the different activity patterns, but also numerically reproduces the endogenous pseudo-oscillatory network dynamics, and their thermal dependence. The inferred Ising model is then analyzed within the so-called mean-field formulation, in which the coupling and bias parameters are replaced by their

values averaged over the left and right subpopulations. It yields a two-dimensional representation of the network energy landscape where the preferred states and associated activation barriers can be easily evaluated. We show how this combined data-driven and theoretical approach can be applied to analyze the ARTR response to transient visual stimulation. The latter tilts the energy landscape, strongly favoring some states over others.

3.1 Ising model is not trained to reproduce short-term temporal correlations, but is able to predict long-term dynamics

The graphical model we introduced in this work was trained to capture the low-order statistics of snapshots of activity. Because graphical models are blind to the dynamical nature of the population activity, it is generally believed that they cannot reproduce any dynamical feature. Nevertheless, here we demonstrate that our model can quantitatively replicate aspects of the network long-term dynamics such as the slow alternation between the two preferred states. To better understand this apparent paradox, it is necessary to distinguish short and long time scales. At short time scale, defined here as the duration of a time bin (of the order of a few 100 ms), the model cannot capture any meaningful dynamics. The Monte Carlo algorithm we used to generate activity is an abstract and arbitrary process, and the correlations it produces between successive time bins can not reproduce the ones in the recording data. Capturing the short-term dynamics would require a biologically-grounded model of the cell-cell interactions, or, at the very least, to introduce parameters capturing the experimental temporal correlations over this short time scale (Marre et al., 2009; Mézard & Sakellariou, 2011).

Yet, the inability of the Ising model to reproduce short time dynamical correlations does not hinder its capacity to predict long-time behavior. The separation between individual neuronal processes (taking place over time scales smaller than 100 ms) and network-scale activity modulation, which happens on time scales ranging from 1 to 20 s is here essential. The weak dependence of macroscopic processes on microscopic details is in fact well known in many fields outside neuroscience. A classical example is provided by chemical reactions, whose kinetics are often controlled by a slow step due to the formation of the activated complex and to the crossing of the associated energy barrier ΔE, requiring a time proportional to $e^{\Delta E/(kT)}$. All fast processes, whose modelling can be very complex, contribute an effective microscopic time scale τ in Arrhenius' expression for the reaction time, see Eq. 4. In this respect, what really matters to predict long time dynamical properties is a good estimate of ΔE, or, equivalently, of the effective energy landscape felt by the system. This is precisely what the Ising model is capable of doing. This explains why, even if temporal information are not explicitly included in the training process, our model may still be endowed with a predictive power over the long-term network dynamics.

3.2 Energy-landscape-based mechanism for persistence

In a preceding article (Wolf et al., 2017), we developed a mathematical model of the ARTR in which the left and right ARTR population were represented by a single unit. To account for the ARTR persistent dynamics, an intrinsic adaptation time-scale had to be introduced in an ad-hoc fashion. While the mean-field version of the inferred Ising model shows some formal mathematical similarity with this two-unit model, it differs in a fundamental aspect. Here, the slow dynamics reflects the itinerant exploration of a two-dimensional energy landscape (Figure 5C), for which the barriers separating metastable states scale linearly with the system size. The time to cross these barriers in turn grows exponentially with the system size, as prescribed by Arrhenius law, and can be orders of magnitude larger than any single-neuron relaxation time. Persistence is therefore an emerging property of the neural network.

3.3 Mean-field approximation and beyond

The mean-field approach, through a drastic simplification of the Ising model, allows us to unveil the fundamental network features controlling its coarse-grained dynamics. within this approximation, the distributions of couplings and of biases are replaced by their average values. The large heterogeneity that characterizes the Ising model parameters (Suppl. Fig. S1F-I), and is ignored in the mean–field approach, may however play an important role in the network dynamics.

In the Ising model, the ipsilateral couplings are found to be broadly distributed such as to possess both negative and positive values. This leads to the presence of so-called frustrated loops, that is, chains of neurons along which the product of the pairwise couplings is negative. The states of activities of the neurons along such loops cannot be set in a way that satisfies all the excitatory and inhibitory connections, hence giving rise to dynamical instabilities in the states of the neurons. The absence of frustrated loops in the network (Figure 5A) stabilizes and boosts the activity, an artifact we had to correct for in our analytical treatment by introducing an effective number of neurons K, much smaller than the total numbers of neurons Ns. Neglecting the variability of the contralateral couplings also constitutes a drastic approximation of the mean field approach. This is all the more true that the average contralateral coupling I happens to be small compared to its standard deviation.

Couplings are not only broadly distributed but also spatially organized. Ipsilateral couplings J_{ij} decay with the distance between neurons i and j (Suppl. Fig. S1J). Similarly, contralateral couplings show strong correlations for short distances between the contralateral neurons (Suppl. Fig. S1K). The existence of a local spatial organization in the couplings is not unheard of in computational neuroscience, and can have important functional consequences. it is for instance at the basis of ring-like attractor models and their extensions to 2 or 3 dimensions (Tsodyks & Sejnowski, 1995). Combined

with the presence of variable biases h_i, short-range interactions can lead to complex propagation phenomena, intensively studied in statistical physics in the context of the Random Field Ising Model. (Schneider & Pytte, 1977; Kaufman et al., 1986). As the most excitable neurons (with the largest biases) fire they excite their neighbors, who in turn become active, triggering the activation of other neurons in their neighborhood. Such an avalanche mechanism could explain the fast rise of activity in the left or right region, from low- to high-activity state.

3.4 Interpretation of the functional connectivity

The inferred functional couplings J_{ij}'s are not expected to directly reflect the corresponding structural (synaptic) connectivity. However, their spatial distribution appears to be in line with the known ARTR organization (Dunn et al., 2016; Kinkhabwala et al., 2011) characterized by large positive (excitatory) interactions within the left and right population, and by the presence of negative (inhibitory) contralateral interactions. Although the contralateral couplings are found to be, on average, almost null, compared to the ipsilateral excitatory counterparts, they drive a subtle interplay between the left and right regions of the ARTR.

As shown in Suppl. Table S2 the inferred parameters largely vary across datasets. This variability is partially due to the difficulty to separately infer the interactions J_{ij} and the biases h_i, a phenomenon not specific to graphical model but also found with other neural *e.g.* Integrate-and-Fire network models (Monasson & Cocco, 2011). This issue can be easily understood within mean-field theory. For simplicity let us neglect the weak contralateral coupling I. The mean activity m of a neuron then depends on the total 'input' $J\,m + H$ it receives, which is the sum of the bias H and of the mean ipsilateral activity m, weighted by the recurrent coupling J. Hence, the combination $J\,m + H$ is more robustly inferred than H and J taken separately, see Suppl. Fig. S2E-F.

Our neural recordings demonstrate a systematic modulation of the ARTR dynamics with the bath temperature, in quantitative agreement with the thermal-dependance of the exploratory behavior in freely-swimming assays. The model correctly captures this thermal modulation of the ARTR activity, and in particular the decay of the persistence time with the temperature. This owes to a progressive change in the values of both the couplings and the biases, which together deform the energy landscape and modulate the energy barriers between metastable states. Temperature can have direct effects on cellular and synaptic processes, which may explain these observed changes in the functional couplings. Another possibility is that these changes would result from a temperature-dependent release of neuromodulatory signals throughout the brain.

The capacity to quantitatively capture subtle differences in the spontaneous activity induced by external cues is an important asset of our model. Recent studies have shown that spontaneous behavior in zebrafish larvae is not time-invariant but exhibits transitions between different regimes, lasting

over minutes and associated with specific brain-states. These transitions can have no apparent cause (Le Goc et al. (2021)) or be induced by external (*e.g.* stimuli, Andalman et al. (2019)) or internal cues (*e.g.* hunger states, Marques et al. (2019)). Although they engage brain-wide changes in the pattern of spontaneous neural dynamics, they are often triggered by the activation of neuromodulatory centers such as the habenula-dorsal raphe nucleus circuit (Corradi & Filosa, 2021). Training Ising models in various conditions may help decipher how such neuromodulation impacts the network functional couplings leading to distinct dynamical regimes of spontaneous activity.

3.5 Data-driven modelling and metastability

With its slow alternating activity and relatively simple architecture, the ARTR offers an ideally suited circuit to test the capacity of Ising models to capture network-driven dynamics. The possibility to experimentally modulate the ARTR persistence time-scale further enabled us to evaluate the model ability to quantitatively represent this slow process. The ARTR is part of a widely distributed hindbrain network that controls the eye horizontal saccadic movements, and which includes several other neuronal populations whose activity is tuned to the eye velocity or position (Joshua & Lisberger, 2015; Wolf et al., 2017). A possible extension of the model would consist in incorporating these nuclei in order to obtain a more complete representation of the oculomotor circuit. Beyond this particular functional network, a similar data-driven approach could be implemented to capture the slow concerted dynamics that characterize numerous neural assemblies in the zebrafish brain (van der Plas et al., 2021).

The importance of metastable states in cortical activity in mammals has been emphasized in previous studies as a possible basis for sequence-based computation (Harvey et al., 2012; Brinkman et al., 2021). Our model suggests that these metastable states are shaped by the connectivity of the network, and are naturally explored during ongoing spontaneous activity. In this respect, the modification of the landscape resulting from visual stimulation, leading to a sharp decrease in the barrier separating the states is reminiscent of the acceleration of sensory coding reported in Mazzucato et al. (2019). Our principled data-driven modeling could be useful to assess the generality of such metastable-state-based computations and of their modulation by sensory inputs in other organisms.

4 Acknowledgements

S.W. acknowledges support by a fellowship from the Fondation pour la Recherche Médicale (SPF 201809007064), and G.L.G. by the Systems Biology network of Sorbonne Université. This project was in part funded by the Human Frontier Science Program (RGP0060/2017).

We thank the IBPS fish facility staff for the fish maintenance, we are grateful to Carounagarane Dore for his contribution to the design of the experimental setups. We thank Misha Ahrens for providing the GCaMP line.

5 Author Contributions

Functional neural recordings and behavioral experiments were carried out and analyzed by G.L.G. and G.D.; S.W., S.C. and R.M. inferred Ising models from activity recordings, performed Monte Carlo simulations, and derived the mean-field analysis. The project was conceived and supervised by R.M, S.C and G.D. All authors contributed to the writing and editing of the manuscript.

6 Declaration of Interests

The authors declare no competing interests.

7 STAR Methods

7.1 Zebrafish lines and maintenance

All animals subjects were zebrafish (*Danio rerio*), aged 5 to 7 days post-fertilization (dpf). Larvae were reared in Petri dishes in embryo medium (E3) on a 14/10h light/dark cycle at 28°C, and were fed powdered nursery food (GM75) every day from 6dpf.

Calcium imaging experiments were conducted on *nacre* mutants that were expressing either the calcium indicator GCaMP6f (12 fish) or GCaMP6s (1 fish) in the nucleus under the control of the nearly pan-neuronal promoter *Tg(elavl3:H2B-GCaMP6)*. Both lines were provided by Misha Ahrens and published in Vladimirov et al. (2014) (H2B-GCaMP6s) and Quirin et al. (2016) (H2B-GCaMP6f).

All experiments were approved by Le Comité d'Éthique pour l'Expérimentation Animale Charles Darwin (02601.01).

7.2 Behavioral assays

The behavioral experiments and pre-processing have been described in details elsewhere (Le Goc et al., 2021). Shortly, it consists in a metallic pool regulated in temperature with two Peltier elements, recorded in uniform white light from above at 25Hz. Batch of 10 animals experienced 30min in water at either 18, 22, 26, 30 or 33°C (10 batches of 10 fish, involving 170 different individuals, were used). Movies were tracked with FastTrack (Gallois & Candelier, 2021), and MATLAB (The Mathworks) is

used to detect discrete swim bouts from which the differences of orientation between two consecutive events are computed, referred to as turn or reorientation angles $\delta\theta$.

Turn angles distributions could be fitted as the sum of two distributions (Gaussian and Gamma), whose intersection was used to define an angular threshold to categorize events into forward (F), left turn (L) or right turn (R, Figure 1A). This threshold was found to be close to 10 degrees for all tested temperatures.

Then we ternarized $\delta\theta$ values, based on F, L or R classification (Figure 1B) and computed the power spectrum of the binary signals defined from symbols L and R only, with the `periodogram` MATLAB function and averaged by temperature (Figure 1C). The outcome was fitted to the Lorentzian expression corresponding to a memory-less equiprobable two-state process (Odde & Buettner, 1998):

$$S\left(f\right) \propto \frac{2k_{flip}}{4k_{flip}^2 + \left(2\pi f\right)^2} \ , \tag{5}$$

where k_{flip} is the rate of transition from one state to another. The inverse of the fitted flipping rate k_{flip} represents the typical time spent in the same orientational state, *i.e.* the typical time taken to switch turning direction.

7.3 Light-sheet functional imaging of spontaneous activity

Volumetric functional recordings were carried out using custom-made one-photon light-sheet microscopes whose optical characteristics have been detailed elsewhere (Panier et al., 2013). Larvae were mounted in a 1mm diameter cylinder of low melting point agarose at 2% concentration.

Imaged volume corresponded to 122 ± 46 μm in thickness, split into 16 ± 4 slices (mean ± s.d.). Recordings were of length 1392 ± 256 seconds with a brain volume imaging frequency of 6 ± 2 Hz (mean ± s.d.).

Image pre-processing, neurons segmentation and calcium transient ($\Delta F/F$) extraction were performed offline using MATLAB, according to the workflow previously reported (Panier et al., 2013; Wolf et al., 2017; Migault et al., 2018).

A Peltier module is attached to the lower part of the pool (made of tin) with thermal tape (3M). A type T thermocouple (Omega) is placed near the fish head (< 5mm) to record the fish surrounding temperature. The signal from a thermocouple amplifier (Adafruit) is used in a PID loop implemented on an Arduino board, which mitigate the Peltier power to achieve the predefined temperature target, stable at $\pm 0.5°C$. The temperature regulation softwares and electronics design are available on Gitlab under a GNU GPLv3 licence (`https://gitlab.com/GuillaumeLeGoc/arduino-temperature-control`).

The ARTR neurons were selected using a method described elsewhere (Wolf et al., 2017). First, a group of neurons was manually selected on a given slice based on a morphological criterion such that the ARTR structure (ipsilateral correlations and contralateral anticorrelation) is revealed. Then,

neurons showing Pearson's correlation (anti-correlation) higher than 0.2 (less than -0.15, respectively) are selected, manually filtering them on a morphological criterion. Those neurons are then added to the previous ones, whose signals are used to find neurons from the next slice and so on until all slices are treated.

For fish that were recorded at different temperature, to ensure that the same neurons are selected, we used the Computational Morphometry Toolkit (CMTK, `https://www.nitrc.org/projects/cmtk/`) to align following recordings onto the first one corresponding to the same individual. Resulting transformations are then applied to convert neurons coordinates in a consistent manner through all recordings involving the same fish.

7.4 Time constants definitions

For the flipping rates (Figure 1J), we defined the time-dependent signed activity of the ARTR (Figure 1H) through

$$\sigma(t) = \mathrm{sign}\big(m_L(t) - m_R(t)\big) \ , \tag{6}$$

where $m_{L,R}(t) = \frac{1}{N_{L,R}} \sum_{i \in L,R} s_i(t)$ are the average activities in the L,R regions. A power spectrum density is estimated for each signal with the Thomson's multitaper method through the `pmtm` MATLAB function (time-halfbandwidth product set to 4). The power spectrum densities were then fitted with a Lorentzian spectrum, see Eq. 5.

ARTR left and right persistence times (Figure 2E) are defined as the time m_L and m_R signals spend consecutively above an arbitrary threshold set at 0.1. Left and right signals are treated altogether. Changing the threshold does induce a global offset but does not change the observed effect of temperature, the relation with m_L and m_R mean signals, nor the relation with the persistence times of the synthetic signals. The latter are computed according to the same procedure.

7.5 Visually-driven recordings

Volumetric functional recordings under visual stimulation were carried using our two-photon light-sheet microscope described in Wolf et al. (2015). The stimulation protocol was previously explained in Wolf et al. (2017): two LEDs were positioned symmetrically outside of the chamber at 45° and 4.5 cm from the fish eyes, delivering a visual intensity of 20 μW/cm2. We alternately illuminated 17 times each eye for 10s, 15s, 20s, 25s and 30s while performing two-photon light-sheet brain-wide functional imaging. Synchronization between the microscope and the stimulation set-up was done using a D/A card (NI USB-6259 NCS, National Instruments) and a LabVIEW program. Brain volume image frequency was of 1Hz on the 7 recorded fish. Recordings last for 4500s, 856s of wich is spontaneous activity. We extracted the ARTR neurons following the same procedure described above, yielding 89 $\pm$ 54 neurons (mean $\pm$ s.d.).

7.6 Inference of Ising model from neural activity

7.6.1 From spontaneous activity to spiking data, to biases and connectivity

For each recording (animal and/or temperature) the binarized spike trains were inferred from the fluorescence activity signal using the Blind Sparse Deconvolution algorithm (Tubiana et al., 2020). The binarized activity of the N recorded neurons was then described for each time bin t, into a N-bit binary configuration $\mathbf{s}_t$, with , $s_i\,(t) = 1$ if neuron i is active in bin t, 0 otherwise.

The functional connectivity matrix $J_{\rm ij}$ and the biases h_i defining the Ising probability distribution over neural configurations, see Eq. 1, were determined such that the pairwise correlations and average activities computed from the model match their experimental counterparts. In practice, we approximately solved this hard inverse problem using the Adaptative Cluster Expansion and the Monte-Carlo learning algorithms described in Cocco & Monasson (2011) and in Barton & Cocco (2013). The full code of the algorithms can be downloaded from the GitHub repository : `https://github.com/johnbarton/ACE/`.

7.6.2 Inference of additional biases from visually-driven activity recordings

For the visually-driven activity recordings, we infer the additional biases $\delta h_i^{\leftarrow}$ from the recordings of the ARTR activity during, for example, the leftward light stimulations as follows. Let $\overleftarrow{B}$ the number of time bins $t = 1, 2, ..., \overleftarrow{B}$ in the recording, and $\mathbf{s}_t$ the corresponding binarized activity configurations. We define, for each neuron i,

$$\rho_i(\delta h) = \sum_{t=1}^{\overleftarrow{B}} \frac{\exp\left(h_i + \sum_j J_{ij}\, s_j(t) + \delta h\right)}{1 + \exp\left(h_i + \sum_j J_{ij}\, s_j(t) + \delta h\right)}\ . \tag{7}$$

$\rho_i(\delta h)$ represents the mean activity of neuron i, when subject to a global bias summing h_i, the other neurons activities $s_j(t)$ weighted by the couplings J_{ij}, and an additional bias δh, averaged over all the frames t corresponding to left-sided light stimulation. It is a monotonously increasing function of δh, which matches the experimental average activity $\frac{1}{\overleftarrow{B}} \sum_{t=1}^{\overleftarrow{B}} s_i(t)$ for a unique value of its argument. This value defines $\delta h_i^{\leftarrow}$.

The same procedure was followed to infer the additional biases $\delta h_i^{\rightarrow}$ associated to rightward visual stimulations.

7.6.3 Real data and Ising models comparison

To quantify the quality of the log-probability landscapes reproduction by the Ising models (Figure 4C), we used the Kullback-Leibler divergence between (1) a dataset i and the synthetic signals generated with the model trained on that dataset i (green) and (2) the dataset i with synthetic signals generated with every other models (red). With c_i the count in the two-dimensional bin i (10×10

bins used) and α a pseudocount (set to 1), the probability in bin i is defined as $P_i = \frac{c_i+\alpha}{\sum_j (c_j+\alpha)}$. The Kullback-Leibler divergence between a data/model pair is then defined as

$$D_{KL} = \sum_i P_{data,i} \log_{10} \left(\frac{P_{data,i}}{P_{model,i}} \right) \tag{8}$$

7.7 Mean-field theory for the ARTR activity

7.7.1 Derivation of the free energy

We consider an Ising model with N_L and N_R neurons in, respectively, the left and right regions. Each neuron activity variable can take two values, $\sigma_i = 0, 1$, corresponding to silent and active states (within a time window). The "energy" of the system reads

$$E(s_1, \ldots, s_{N_L}, s_{N_L+1}, \ldots, s_{N_L+N_R}) = -\tilde{H}_L \sum_{i=1}^{N_L} s_i - \tilde{H}_R \sum_{i=N_L+1}^{N_L+N_R} s_i - \frac{1}{2} \sum_{i \neq j} \tilde{J}_{ij}\, s_i\, s_j \ , \tag{9}$$

where $\tilde{H}_L, \tilde{H}_R$ are biases acting on the neurons, and the coupling matrix is defined through

$$\tilde{J}_{ij} = \begin{cases} \tilde{J}_L & \text{if} \quad 1 \leq i,j \leq N_L \ , \\ \tilde{J}_R & \text{if} \quad N_L+1 \leq i,j \leq N_L+N_R \ , \\ \tilde{I} & \text{otherwise} \ . \end{cases} \tag{10}$$

We now introduce the left and right average activities:

$$m_L = \frac{1}{N_L} \sum_{i=1}^{N_L} s_i \quad , \qquad m_R = \frac{1}{N_R} \sum_{i=N_L+1}^{N_L+N_R} s_i \ . \tag{11}$$

The energy E of a neural activity configuration in Eq. 9 can be expressed in terms of these average activities:

$$\begin{aligned} E(m_L, m_R) \quad &= \quad -N_L \left(\tilde{H}_L - \frac{\tilde{J}_L}{2} \right) m_L - N_R \left(\tilde{H}_R - \frac{\tilde{J}_R}{2} \right) m_R \\ &- \quad \frac{(N_L)^2}{2} \tilde{J}_L\, m_L^2 - \frac{(N_R)^2}{2} \tilde{J}_R\, m_R^2 - \tilde{I}\, N_L\, N_R\, m_L\, m_R \ . \end{aligned} \tag{12}$$

We may now compute the partition function normalizing the probability of configurations, see Eq. 1,

$$Z = \sum_{\{s_i=0,1\}} e^{-E(s_1,\ldots,s_{N_L+N_R})} = \sum_{m_L, m_R} \mathcal{M}_L(m_L)\, \mathcal{M}_R(m_R)\, e^{-E(m_L,m_R)} \ , \tag{13}$$

where the sums runs over fractional values of the average left and right activities, from 0 to 1 with steps equal to, respectively, $2/N_L$ and $2/N_R$, and the multiplicities $\mathcal{M}_L$ and $\mathcal{M}_R$ measure the numbers of neural configurations with prescribed average activities. We approximate these multiplicities with the standard entropy-based expressions, which are exact in the limit of large sizes K_L, K_R:

$$\mathcal{M}_L(m_L) \simeq e^{N_L\, S(m_L)} \quad , \qquad \mathcal{M}_R(m_R) \simeq e^{N_R\, S(m_R)} \ , \tag{14}$$

where

$$S(m) = -m \ln m - (1-m) \ln(1-m) \tag{15}$$

is the entropy of a $0-1$ variable with mean m. As a consequence the activity-dependent free energy $F(m_L, m_R)$ is given by

$$\begin{aligned} F(m_L, m_R) &= E(m_L, m_R) - N_L\, S(m_L) - N_R\, S(m_R) \\ &= -\frac{N_L\, J_L}{2}\, {m_L}^2 - \frac{N_R\, J_R}{2}\, {m_R}^2 - I\, \sqrt{N_L\, N_R}\, m_L\, m_R - N_L\, H_L\, m_L - N_R\, H_R\, m_R \\ &+ N_L\, (m_L\, \ln m_L + (1-m_L)\, \ln(1-m_L)) + N_R\, (m_R\, \ln m_R + (1-m_R)\, \ln(1-m_R)) \end{aligned} \tag{16}$$

where the bias and coupling parameters are, respectively, $H_L = \tilde{H}_L - \frac{\tilde{J}_L}{2}$, $H_R = \tilde{H}_R - \frac{\tilde{J}_R}{2}$, $J_L = N_L\, \tilde{J}_L$, $J_R = N_R\, \tilde{J}_R$, $I = \sqrt{N_L\, N_R}\, \tilde{I}$.

The sizes N_L, N_R enter formula (16) for the free energy in two ways:

- *implicitly*, through the biases H_L, H_R and the couplings J_L, J_R, I. These parameters are equal to, respectively, the average bias and the total ipsilateral and contralateral couplings acting on each neuron in the L and R regions. They are effective parameters defining the mean-field theory;
- *explicitly*, as multiplicative factors to the free energy contributions coming from the left and right regions. The sizes then merely act as effective inverse "temperatures", in the Boltzmann factor $e^{-F(m_L, m_R)}$ associated to the probability of the L, R activities.

Mean-field theory generally overestimates the collective effects of interactions; a well-known illustration of this artifact is the prediction of the existence of a phase transition in the uni-dimensional ferromagnetic Ising model with short range interactions, while such a transition is rigorously known not to take place (Ma, 1985). We expect these effects to be strong here, due to the wide distribution of inferred Ising couplings (Figure 3C and Suppl. Fig. S1I). Many pairs of neurons carry close to zero couplings, and the interaction neighborhood of a neuron is effectively much smaller than N_L and N_R. To compensate for the overestimation of interaction effects we thus propose to keep Eq. 16 for the free energy, but with effective sizes K_L, K_R replacing the numbers N_L, N_R of recorded neurons, see Eq. 2, leading to the expression of the free energy:

$$\begin{aligned} F\,(m_L, m_R) &= -\frac{K_L\, J_L}{2}\, {m_L}^2 - \frac{K_R\, J_R}{2}\, {m_R}^2 - I\, \sqrt{K_L\, K_R}\, m_L\, m_R - K_L\, H_L\, m_L - K_R\, H_R\, m_R \\ &+ K_L\, (m_L\, \ln m_L + (1-m_L)\, \ln(1-m_L)) + K_R\, (m_R\, \ln m_R + (1-m_R)\, \ln(1-m_R)) \end{aligned} \tag{17}$$

These effective sizes K_L, K_R are expected to be smaller than N_L, N_R, *i.e.* the associated "temperatures" (inverse sizes) are expected to be higher. Their values are fixed through the comparison of the Langevin dynamical traces with the traces coming from the data, see below.

7.7.2 Langevin dynamical equations

The dynamical Langevin equations (3) read

$$\tau \frac{dm_L}{dt} = K_L\big(J_L\, m_L + H_L\big) + I\sqrt{K_L\, K_R} m_R - K_L\, \log\left(\frac{m_L}{1-m_L}\right) + \epsilon_L(t)\ , \tag{18}$$

$$\tau \frac{dm_R}{dt} = K_R\big(J_R\, m_R + H_R\big) + I\sqrt{K_L\, K_R} m_L - K_R\, \log\left(\frac{m_R}{1-m_R}\right) + \epsilon_R(t)\ , \tag{19}$$

where ϵ_L, ϵ_R denote white-noise processes, see main text.

7.7.3 Fit of the effective sizes K_L and K_R

The effective sizes $K_L = N_L/A$ and $K_L = N_R/A$ were fitted generating Langevin trajectories of the activities (m_L,m_R) for a large set of values of A (i.e. K_L and K_R), and with fixed parameters (H_L,H_R, J_L,J_R,τ). For each value of K_L and K_R we compute the Kullback-Leibler (KL) divergence between the experimental and the Langevin distributions of (m_L,m_R) (see Suppl. Fig. S2A-C). The effective sizes K_L and K_R are the ones that minimize the value of the KL divergence. For low values of A the KL divergence can be noisy and creates artifacts. To avoid these artifacts we assume that $A > 2$.

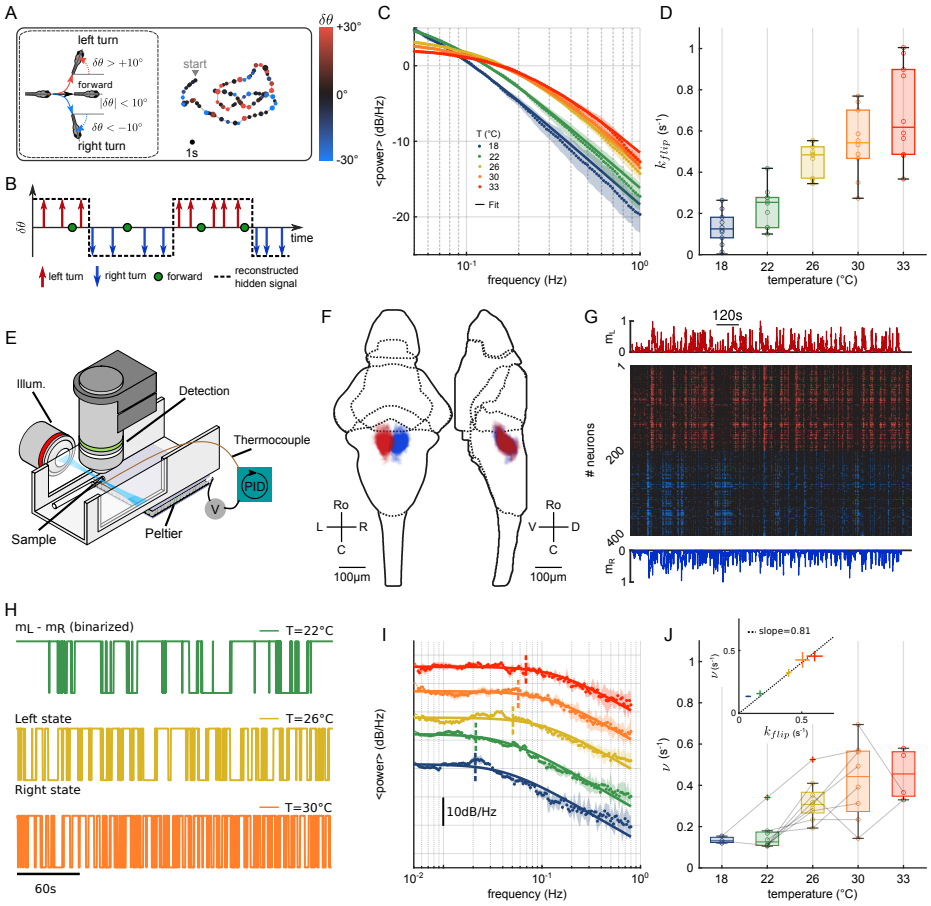

Figure 1: **Temperature-dependence of orientational persistence and ARTR dynamics.**

A, Swimming patterns in zebrafish larvae. Swim bouts are categorized into forward and turn bouts, based on the amplitude of the heading reorientation. Example trajectory: each dot corresponds to a swim bout; the color encodes the reorientation angle. **B**, The bouts are discretized as left/forward/right bouts. The continuous binary signal represents the putative orientational state governing the chaining of the turn bouts. **C**, Power spectra of the discretized turn signal averaged over all animals for each temperature (dots). Each spectrum is fitted by a Lorentzian function (solid lines) from which we extract the switching rate k_{flip}. **D**, Temperature dependence of k_{flip}. **E**, Light-sheet based volumetric recording of neuronal activity. The fish bath is controlled in temperature using a Peltier device and a thermocouple probe. **F**, Morphological organization of the ARTR. **G**, Raster plot of the ARTR spontaneous dynamics showing antiphasic right/left alternation. The top and bottom traces are the ARTR average signal of the left and right subcircuits. **H**, Example ARTR ($m_L - m_R$) binarized signals measured at 3 different temperatures (same larva). **I**, Averaged power spectrum of the ARTR signals, for the 5 tested temperatures. The dotted vertical lines indicate the signal switching frequencies ν as extracted from the Lorentzian fit (solid lines). **J**, Temperature-dependence of ν. The lines join data points obtained with the same larva. Inset: relationship between k_{flip} (behavioral) and ν (neuronal) switching frequencies. The dashed line is the linear fit.

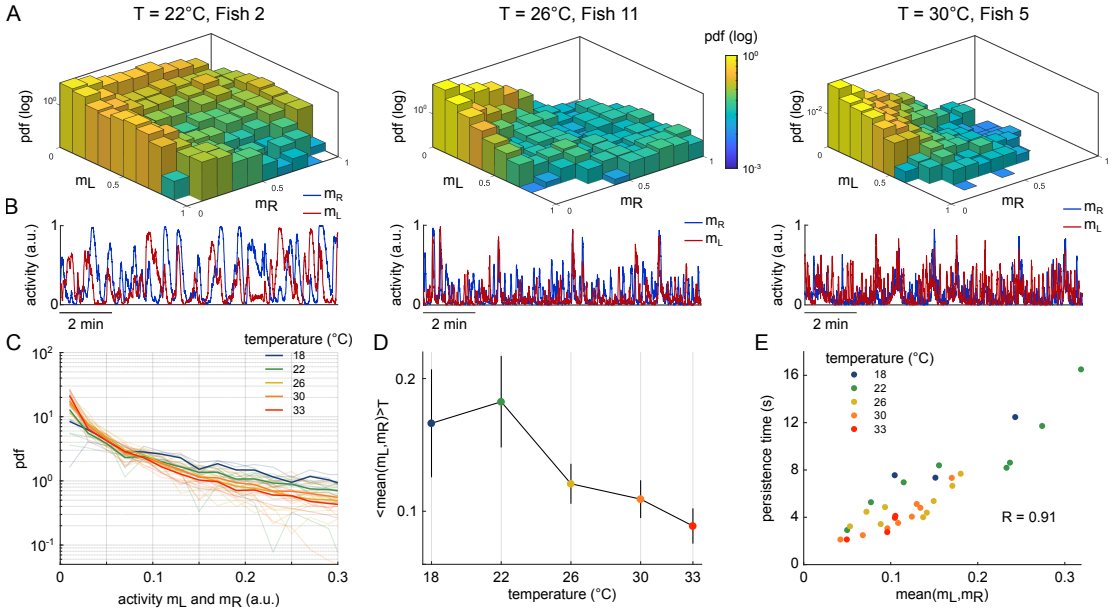

Figure 2: **Temperature-dependence of ARTR state occupancy.**

A, Probability densities $P(m_L, m_R)$, see Eq. 2, of the activity state of the circuit, in logarithmic scale; Color encodes z-axis (common colorbar in the middle panel). Left: 22°C; Middle: 26°C; Right: 30°C; Three different fish. **B**, Corresponding 10 min long time-signals of the mean activities of the left (red) and right (blue) subregions of the ARTR. **C**, Pdf of activities of both sides of the ARTR. Color encodes temperature. **D**, Temperature-averaged mean activity from both sides of the ARTR. Error bars are standard error of the mean. **E**, Duration of sustained activity of both sides of the ARTR (persistence time) vs. mean activity. Each dot is the mean persistence time from one fish at one temperature, colors encode temperature.

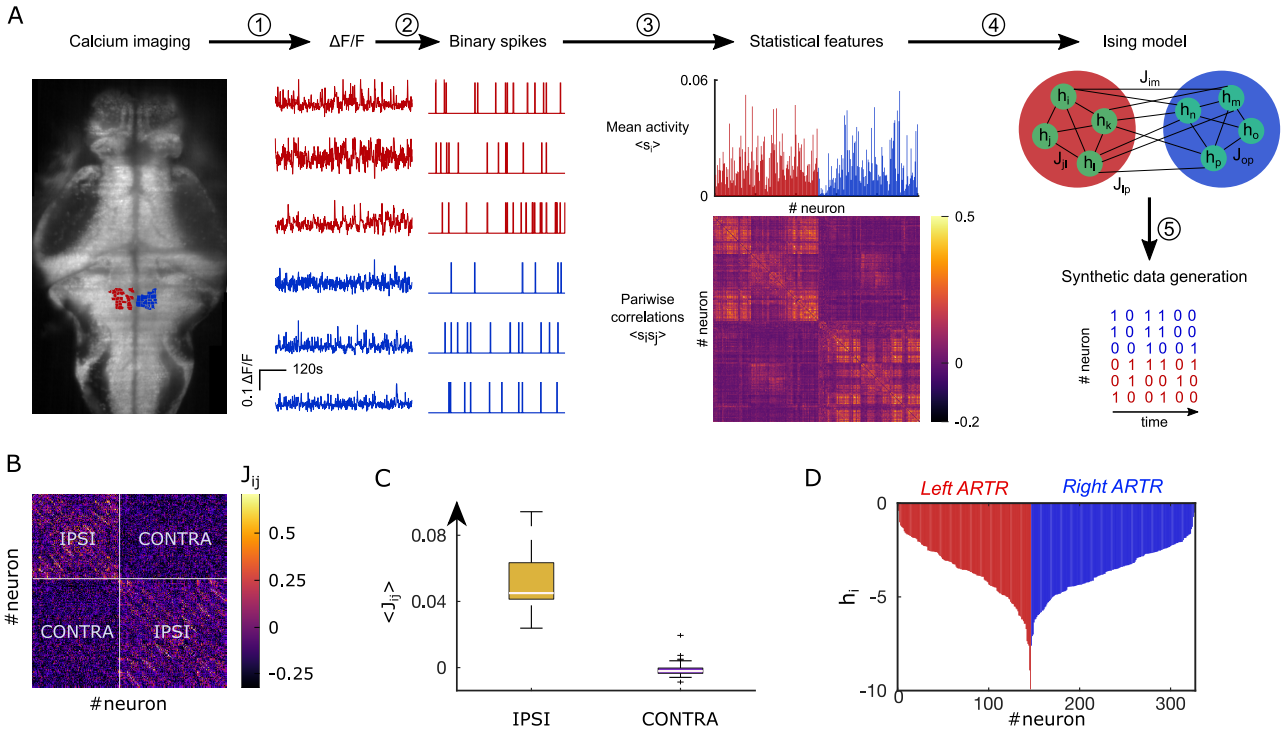

Figure 3: **Ising model inference from the recordings of ARTR activity.**

A, Processing pipeline for the inference of the Ising model. We first extract from the recorded calcium activity, approximate spike trains using a Bayesian deconvolution algorithm (BSD). The activity of each neuron is then "0" or "1". We then compute the mean activity and the pairwise covariance of the data, from which we infer the parameters h_i and J_{ij} of the ising model. Finally, we can generate synthetic data using Monte-Carlo sampling. **B**, Functional connectivity matrix. The white lines separate the left/right part of the ARTR and defines 4 blocks associated with ipsilateral or contralateral couplings. **C**, Box plot across animals of the average value of the ipsilateral and contralateral couplings. **D**, Bias parameter distribution for an example fish.

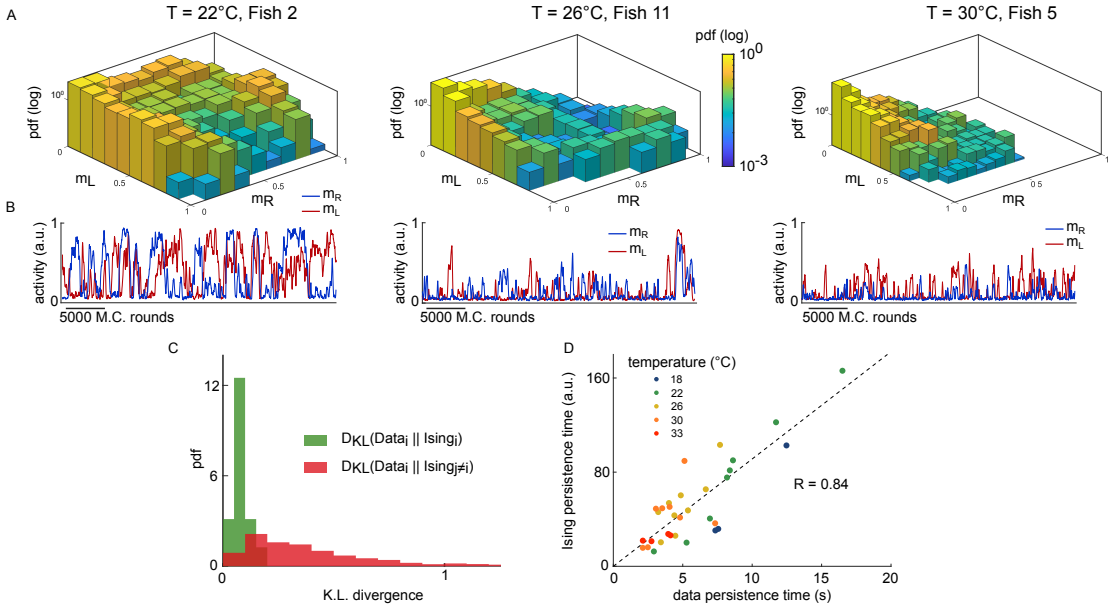

Figure 4: **Ising models reproduce characteristic features of the recorded activity.**

A-B, Three example activity maps and corresponding ARTR signals from simulations trained on the same datasets as in figure 2A-B. **A** Probability distribution of finding the network in the (m_L, m_R) state, in logarithmic scale, color encodes z-axis (common colorbar in the middle panel). **B**, synthetic (MC) signals of the Ising model dynamics, showing persistent activity. The red and blue traces correspond to the mean activity of the left and right subpopulations, respectively. The time is measured in MC rounds. **C**, Distribution of the Kullback-Leibler divergences between datasets and their corresponding Ising models (green) and between datasets and Ising models trained on a different datasets (red). **D**, Average persistence times in simulations vs. experiments. Each dot refers to one fish at one temperature, colors encode temperature.

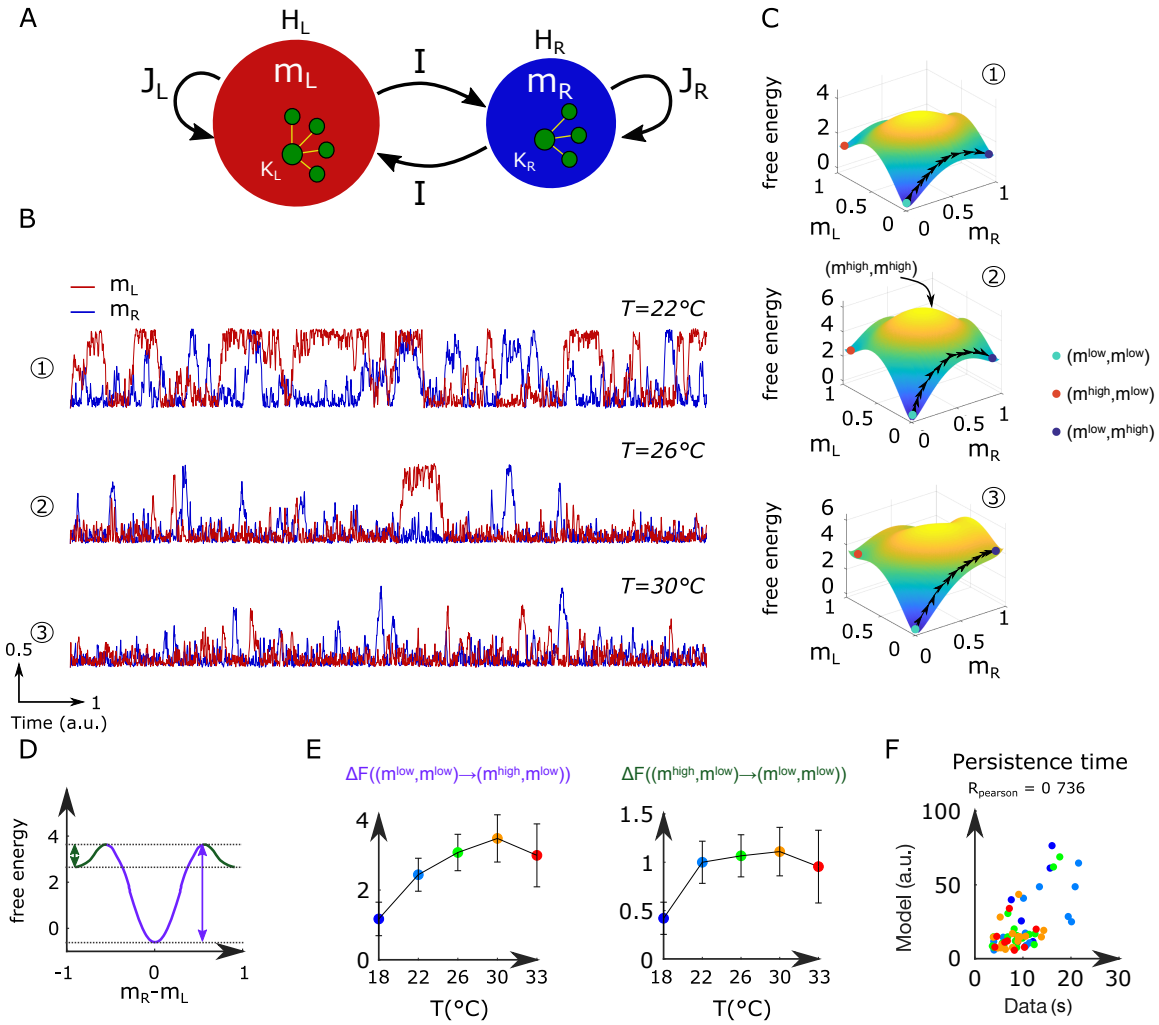

Figure 5: **Mean-field approximation of the inferred Ising model.**

A, Schematic view of the mean-field Ising model. **B**, Examples of simulated m_L and m_R signals of the mean-field dynamical equations for three sets of parameters corresponding to three bath temperatures. **C**, Free-energy landscapes computed with the mean-field model, in the (m_L,m_R) plane for three different parameter sets corresponding to the points 1-3 in panel B. Colored circles denote metastable states, and the line of black arrows represent the optimal path between (m^{low}, m^{low}) and (m^{low}, m^{high}) states. **D**, Schematic free energy profiles along an optimal path, as a function of $m_R - m_L$. The arrows denotes the energy barriers ΔF associated with the various transitions. The dark green arrow denotes $\Delta F\left((m^{high}, m^{low}) \to (m^{low}, m^{low})\right)$, purple denotes $\Delta F\left((m^{low}, m^{low}) \to (m^{high}, m^{low})\right)$. **E**, Values of the free-energy barriers as a function of temperature. Error bars are standard error of the mean. **F**, Persistence time of the mean-field ARTR model for all fish and runs at different experimental temperatures. Each dot refers to one fish at one temperature, colors encode temperature.

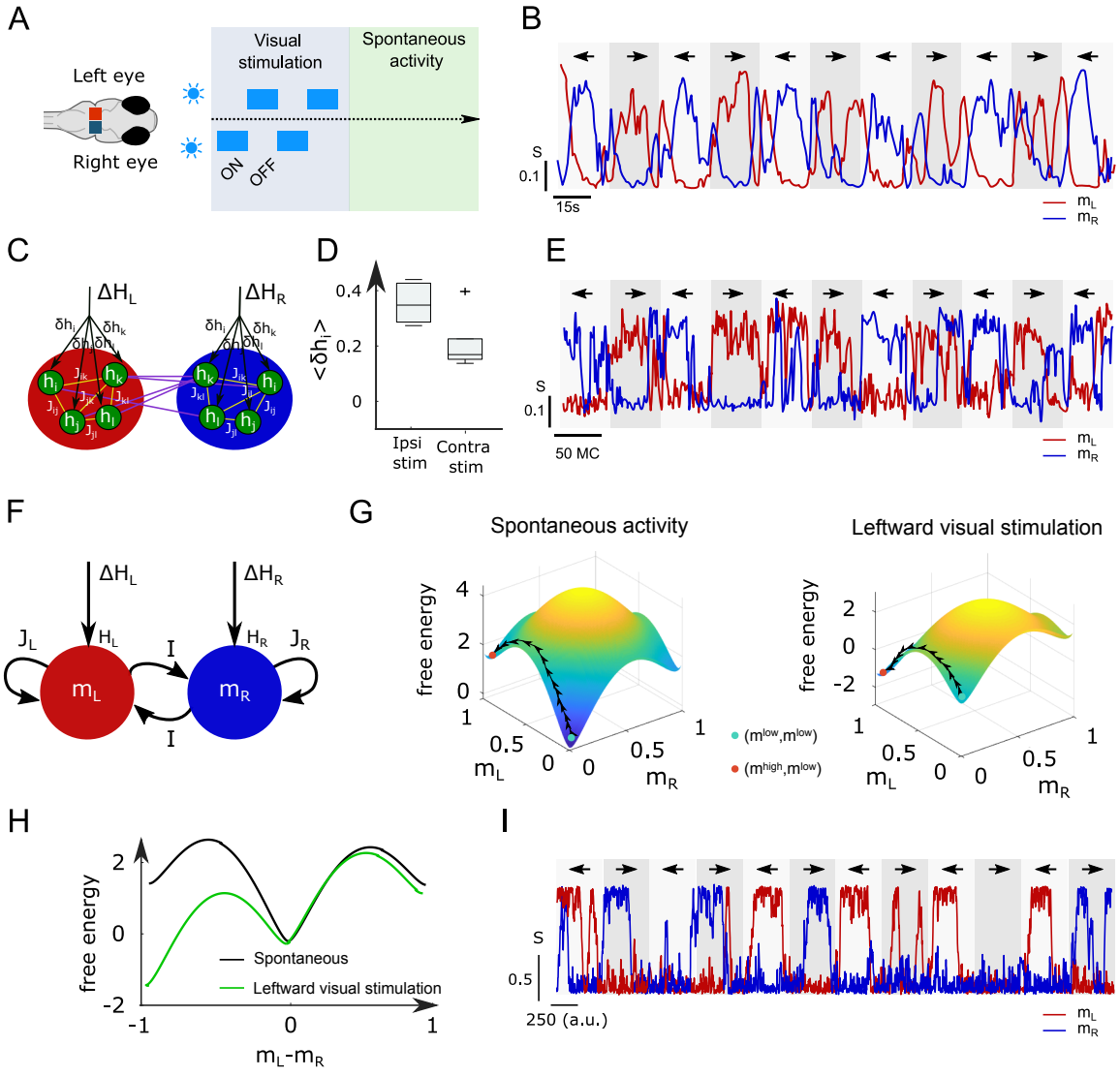

Figure 6: **Modified Ising model captures the behavior of ARTR under visual stimulation.**

A Scheme of the stimulation protocol. The left and right eyes are stimulated alternatively for periods of 15 to 30s, after which a period of spontaneous (unstimulated) activity is acquired. **B**, Example of ARTR activity signals under alternated left-right visual stimulation. The small arrows indicate the direction of the stimulus. **C**, Sketch of the modified Ising model, with additional biases δh_i to account for the local visual inputs. **D**, Values of the additional biases averaged over the ipsilateral and contralateral (with respect to the stimulated eye) neural populations. **E**, Monte Carlo activity traces generated with the modified Ising model. **F**, Sketch of the modified mean-field approach in the presence of visual stimulation. **G**, Free-energy landscapes computed with the mean-field theory during spontaneous (left panel) and stimulated (right panel) activity for an axample fish. **H**, Free-energy along the optimal path as a function of $m_L - m_R$ during spontaneous (black) and stimulated (green) activity corresponding to panel G. **I**, Example of simulated m_L and m_R signals of the mean-field dynamical Langevin equation, Eq. 3 under alternating left and right visual stimulation.

References

Ahrens, M. B., Orger, M. B., Robson, D. N., Li, J. M., & Keller, P. J. (2013). Whole-brain functional imaging at cellular resolution using light-sheet microscopy. *Nature Methods*, *10*, 413–420. doi:10.1038/nmeth.2434.

Andalman, A. S., Burns, V. M., Lovett-Barron, M., Broxton, M., Poole, B., Yang, S. J., Grosenick, L., Lerner, T. N., Chen, R., Benster, T., Mourrain, P., Levoy, M., Rajan, K., & Deisseroth, K. (2019). Neuronal dynamics regulating brain and behavioral state transitions. *Cell*, *177*, 970–985.e20. doi:10.1016/j.cell.2019.02.037.

Barton, J., & Cocco, S. (2013). Ising models for neural activity inferred via selective cluster expansion: Structural and coding properties. *Journal of Statistical Mechanics: Theory and Experiment*, *2013*, P03002. doi:10.1088/1742-5468/2013/03/P03002.

Barton, J., De Leonardis, E., Coucke, A., & Cocco, S. (2016). Ace: adaptive cluster expansion for maximum entropy graphical model inference. *Bioinformatics*, *32*, 3089–3097.

Brinkman, B. A. W., Yan, H., Maffei, A., Park, I. M., Fontanini, A., Wang, J., & Camera, G. L. (2021). Metastable dynamics of neural circuits and networks. arXiv:2110.03025.

Butts, D. (2019). Data-driven approaches to understanding visual neuron activity. *Annual Review of Vision Science*, *5*, 451–477. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85072265786&doi=10.1146%2fannurev-vision-091718-014731&partnerID=40&md5=89e0759aa1cbb626290451127269aeea. doi:10.1146/annurev-vision-091718-014731.

Chen, X., & Engert, F. (2014). Navigational strategies underlying phototaxis in larval zebrafish. *Frontiers in Systems Neuroscience*, *8*. doi:10.3389/fnsys.2014.00039.

Chen, X., Randi, F., Leifer, A. M., & Bialek, W. (2019). Searching for collective behavior in a small brain. *Phys. Rev. E*, *99*, 052418. URL: https://link.aps.org/doi/10.1103/PhysRevE.99.052418. doi:10.1103/PhysRevE.99.052418.

Cocco, S., Leibler, S., & Monasson, R. (2009). Neuronal couplings between retinal ganglion cells inferred by efficient inverse statistical physics methods. *Proceedings of the National Academy of Sciences*, *106*, 14058–14062. URL: https://www.pnas.org/content/106/33/14058. doi:10.1073/pnas.0906705106. Publisher: National Academy of Sciences _eprint: https://www.pnas.org/content/106/33/14058.full.pdf.

Cocco, S., & Monasson, R. (2011). Adaptive Cluster Expansion for Inferring Boltzmann Machines with Noisy Data. *Physical Review Letters*, *106*, 090601. doi:10.1103/PhysRevLett.106.090601.

Corradi, L., & Filosa, A. (2021). Neuromodulation and behavioral flexibility in larval zebrafish: From neurotransmitters to circuits. *Front. Mol. Neurosci.*, *14*, 718951.

Dunn, T. W., Mu, Y., Narayan, S., Randlett, O., Naumann, E. A., Yang, C.-T., Schier, A. F., Freeman, J., Engert, F., & Ahrens, M. B. (2016). Brain-wide mapping of neural activity controlling zebrafish exploratory locomotion. *eLife*, *5*, e12741. doi:10.7554/eLife.12741.

Engeszer, R. E., Patterson, L. B., Rao, A. A., & Parichy, D. M. (2007). Zebrafish in The Wild: A Review of Natural History And New Notes from The Field. *Zebrafish*, *4*, 21–40. doi:10.1089/zeb.2006.9997.

Gallois, B., & Candelier, R. (2021). FastTrack: An open-source software for tracking varying numbers of deformable objects. *PLOS Computational Biology*, *17*, e1008697. doi:10.1371/journal.pcbi.1008697.

Glaser, J. I., Benjamin, A. S., Chowdhury, R. H., Perich, M. G., Miller, L. E., & Kording, K. P. (2020). Machine learning for neural decoding. *eNeuro*, *7*. URL: https://www.eneuro.org/content/7/4/ENEURO.0506-19.2020. doi:10.1523/ENEURO.0506-19.2020. arXiv:https://www.eneuro.org/content/7/4/ENEURO.0506-19.2020.full.pdf.

Guo, Z. V., Inagaki, H. K., Daie, K., Druckmann, S., Gerfen, C. R., & Svoboda, K. (2017). Maintenance of persistent activity in a frontal thalamocortical loop. *Nature*, *545*, 181–186. URL: https://doi.org/10.1038/nature22324. doi:10.1038/nature22324.

Harvey, C. D., Coen, P., & Tank, D. W. (2012). Choice-specific sequences in parietal cortex during a virtual-navigation decision task. *Nature*, *484*, 62–68.

Jaynes, E. T. (1957). Information theory and statistical mechanics. *Phys. Rev.*, *106*, 620–630. URL: https://link.aps.org/doi/10.1103/PhysRev.106.620. doi:10.1103/PhysRev.106.620.

Joshua, M., & Lisberger, S. (2015). A tale of two species: Neural integration in zebrafish and monkeys. *Neuroscience*, *296*, 80–91. URL: https://linkinghub.elsevier.com/retrieve/pii/S0306452214003601. doi:10.1016/j.neuroscience.2014.04.048.

Karpenko, S., Wolf, S., Lafaye, J., Le Goc, G., Panier, T., Bormuth, V., Candelier, R., & Debrégeas, G. (2020). From behavior to circuit modeling of light-seeking navigation in zebrafish larvae. *eLife*, *9*, e52882. doi:10.7554/eLife.52882.

Kaufman, M., Klunzinger, P. E., & Khurana, A. (1986). Multicritical points in an ising random-field model. *Physical Review B*, *34*, 4766.

Kinkhabwala, A., Riley, M., Koyama, M., Monen, J., Satou, C., Kimura, Y., Higashijima, S.-i., & Fetcho, J. (2011). A structural and functional ground plan for neurons in the hindbrain of zebrafish. *Proceedings of the National Academy of Sciences*, *108*, 1164–1169.

Koller, D., & Friedmann, N. (2009). *Probabilistic Graphical Models: Principles and Techniques*. MIT Press.

Langer, J. (1969). Statistical theory of the decay of metastable states. *Annals of Physics*, *54*, 258–275. URL: https://www.sciencedirect.com/science/article/pii/0003491669901535. doi:https://doi.org/10.1016/0003-4916(69)90153-5.

Le Goc, G., Lafaye, J., Karpenko, S., Bormuth, V., Candelier, R., & Debrégeas, G. (2021). Thermal modulation of zebrafish exploratory statistics reveals constraints on individual behavioral variability. *BMC Biology*, *19*, 1–17.

Leyden, C., Brysch, C., & Arrenberg, A. B. (2021). A distributed saccade-associated network encodes high velocity conjugate and monocular eye movements in the zebrafish hindbrain. *Scientific reports*, *11*, 1–17. URL: https://doi.org/10.1038/s41598-021-90315-2. doi:10.1038/s41598-021-90315-2.

Ma, S.-K. (1985). *Statistical Mechanics*. World Scientific.

Marques, J. C., Li, M., Schaak, D., Robson, D. N., & Li, J. M. (2019). Internal state dynamics shape brainwide activity and foraging behaviour. *Nature*, *577*, 239–243. URL: https://doi.org/10.1038%2Fs41586-019-1858-z. doi:10.1038/s41586-019-1858-z.

Marre, O., El Boustani, S., Frégnac, Y., & Destexhe, A. (2009). Prediction of spatiotemporal patterns of neural activity from pairwise correlations. *Phys. Rev. Lett.*, *102*, 138101. URL: https://link.aps.org/doi/10.1103/PhysRevLett.102.138101. doi:10.1103/PhysRevLett.102.138101.

Mazzucato, L., La Camera, G., & Fontanini, A. (2019). Expectation-induced modulation of metastable activity underlies faster coding of sensory stimuli. *Nat. Neurosci.*, *22*, 787–96.

Meshulam, L., Gauthier, J. L., Brody, C. D., Tank, D. W., & Bialek, W. (2017). Collective Behavior of Place and Non-place Neurons in the Hippocampal Network. *Neuron*, *96*, 1178–1191.e4. doi:10.1016/j.neuron.2017.10.027.

Mézard, M., & Sakellariou, J. (2011). Exact mean-field inference in asymmetric kinetic ising systems. *Journal of Statistical Mechanics: Theory and Experiment*, (p. L07001).

Migault, G., van der Plas, T. L., Trentesaux, H., Panier, T., Candelier, R., Proville, R., Englitz, B., Debrégeas, G., & Bormuth, V. (2018). Whole-Brain Calcium Imaging during Physiological Vestibular Stimulation in Larval Zebrafish. *Current Biology*, *28*, 3723–3735.e6. doi:10.1016/j.cub.2018.10.017.

Monasson, R., & Cocco, S. (2011). Fast inference of interactions in assemblies of stochastic integrate-and-fire neurons from spike recordings. *J. Comput Neurosci.*, *31*, 199–227.

Nghiem, T.-A., Telenczuk, B., Marre, O., Destexhe, A., & Ferrari, U. (2018). Maximum-entropy models reveal the excitatory and inhibitory correlation structures in cortical neuronal activity. *Phys. Rev. E*, *98*, 012402. URL: https://link.aps.org/doi/10.1103/PhysRevE.98.012402. doi:10.1103/PhysRevE.98.012402.

Odde, D. J., & Buettner, H. M. (1998). Autocorrelation Function and Power Spectrum of Two-State Random Processes Used in Neurite Guidance. *Biophysical Journal*, *75*, 1189–1196. doi:10.1016/S0006-3495(98)74038-X.

Pandarinath, C., O'Shea, D., & Collins, J. e. a. (2018). Inferring single-trial neural population dynamics using sequential auto-encoders. *Nature Methods*, *15*, 808–815.

Panier, T., Romano, S. A., Olive, R., Pietri, T., Sumbre, G., Candelier, R., & Debrégeas, G. (2013). Fast functional imaging of multiple brain regions in intact zebrafish larvae using Selective Plane Illumination Microscopy. *Frontiers in Neural Circuits*, *7*. doi:10.3389/fncir.2013.00065.

van der Plas, T. L., Tubiana, J., Goc, G. L., Migault, G., Kunst, M., Baier, H., Bormuth, V., Englitz, B., & Debrégeas, G. (2021). Compositional restricted boltzmann machines unveil the brain-wide organization of neural assemblies. *Biorxiv*, . URL: https://doi.org/10.1101/2021.11.09.467900. doi:10.1101/2021.11.09.467900.

Posani, L., Cocco, S., Ježek, K., & Monasson, R. (2017). Functional connectivity models for decoding of spatial representations from hippocampal CA1 recordings. *Journal of Computational Neuroscience*, *43*, 17–33. doi:10.1007/s10827-017-0645-9.

Posani, L., Cocco, S., & Monasson, R. (2018). Integration and multiplexing of positional and contextual information by the hippocampal network. *PLoS computational biology*, *14*, e1006320.

Quirin, S., Vladimirov, N., Yang, C.-T., Peterka, D. S., Yuste, R., & B. Ahrens, M. (2016). Calcium imaging of neural circuits with extended depth-of-field light-sheet microscopy. *Optics Letters*, *41*, 855. doi:10.1364/OL.41.000855.

Ramirez, A. D., & Aksay, E. R. (2021). Ramp-to-threshold dynamics in a hindbrain population controls the timing of spontaneous saccades. *Nature communications*, *12*, 1–19.

Schneider, T., & Pytte, E. (1977). Random-field instability of the ferromagnetic state. *Physical Review B*, *15*, 1519.

Schneidman, E., Berry, M. J., Segev, R., & Bialek, W. (2006). Weak pairwise correlations imply strongly correlated network states in a neural population. *Nature*, *440*, 1007–1012. URL: http://www.nature.com/articles/nature04701. doi:10.1038/nature04701.

Seung, H., Lee, D. D., Reis, B. Y., & Tank, D. W. (2000). Stability of the Memory of Eye Position in a Recurrent Network of Conductance-Based Model Neurons. *Neuron*, *26*, 259–271. URL: https://linkinghub.elsevier.com/retrieve/pii/S0896627300811551. doi:10.1016/S0896-6273(00)81155-1.

Seung, H. S. (1996). How the brain keeps the eyes still. *Proceedings of the National Academy of Sciences*, *93*, 13339–13344.

Tavoni, G., Cocco, S., & Monasson, R. (2016). Neural assemblies revealed by inferred connectivity-based models of prefrontal cortex recordings. *Journal of Computational Neuroscience*, *41*, 269–293. doi:10.1007/s10827-016-0617-5.

Tavoni, G., Ferrari, U., Battaglia, F. P., Cocco, S., & Monasson, R. (2017). Functional coupling networks inferred from prefrontal cortex activity show experience-related effective plasticity. *Network Neuroscience*, *1*, 275–301. URL: https://direct.mit.edu/netn/article/1/3/275-301/2196. doi:10.1162/NETN_a_00014.

Tkačik, G., Mora, T., Marre, O., Amodei, D., Palmer, S. E., Berry, M. J., & Bialek, W. (2015). Thermodynamics and signatures of criticality in a network of neurons. *Proceedings of the National Academy of Sciences*, *112*, 11508–11513. URL: https://www.pnas.org/content/112/37/11508. doi:10.1073/pnas.1514188112. arXiv:https://www.pnas.org/content/112/37/11508.full.pdf.

Tsodyks, M., & Sejnowski, T. (1995). Associative memory and hippocampal place cells. *International journal of neural systems*, *6*, 81–86.

Tubiana, J., Wolf, S., Panier, T., & Debregeas, G. (2020). Blind deconvolution for spike inference from fluorescence recordings. *Journal of Neuroscience Methods*, *342*, 108763. doi:10.1016/j.jneumeth.2020.108763.

Vladimirov, N., Mu, Y., Kawashima, T., Bennett, D. V., Yang, C.-T., Looger, L. L., Keller, P. J., Freeman, J., & Ahrens, M. B. (2014). Light-sheet functional imaging in fictively behaving zebrafish. *Nature Methods*, *11*, 883–884. doi:10.1038/nmeth.3040.

Wang, X.-J. (2008). Decision making in recurrent neuronal circuits. *Neuron*, *60*, 215–234. URL: https://doi.org/10.1016/j.neuron.2008.09.034. doi:10.1016/j.neuron.2008.09.034.

Wolf, S., Dubreuil, A. M., Bertoni, T., Böhm, U. L., Bormuth, V., Candelier, R., Karpenko, S., Hildebrand, D. G. C., Bianco, I. H., Monasson, R., & Debrégeas, G. (2017). Sensorimotor computation underlying phototaxis in zebrafish. *Nature Communications*, *8*, 651. doi:10.1038/s41467-017-00310-3.

Wolf, S., Supatto, W., Debrégeas, G., Mahou, P., Kruglik, S. G., Sintes, J.-M., Beaurepaire, E., & Candelier, R. (2015). Whole-brain functional imaging with two-photon light-sheet microscopy. *Nature methods*, *12*, 379–380.

Zaksas, D., & Pasternak, T. (2006). Directional signals in the prefrontal cortex and in area MT during a working memory for visual motion task. *Journal of Neuroscience*, *26*, 11726–11742. URL: https://doi.org/10.1523/jneurosci.3420-06.2006. doi:10.1523/jneurosci.3420-06.2006.

Zylberberg, J., & Strowbridge, B. W. (2017). Mechanisms of persistent activity in cortical circuits: Possible neural substrates for working memory. *Annual Review of Neuroscience*, *40*, 603–627. URL: https://doi.org/10.1146/annurev-neuro-070815-014006. doi:10.1146/annurev-neuro-070815-014006.

Supplementary Information

Emergence of time persistence in an interpretable data-driven neural network model

Sébastien Wolf, Guillaume Le Goc, Simona Cocco, Georges Debrégeas, Rémi Monasson

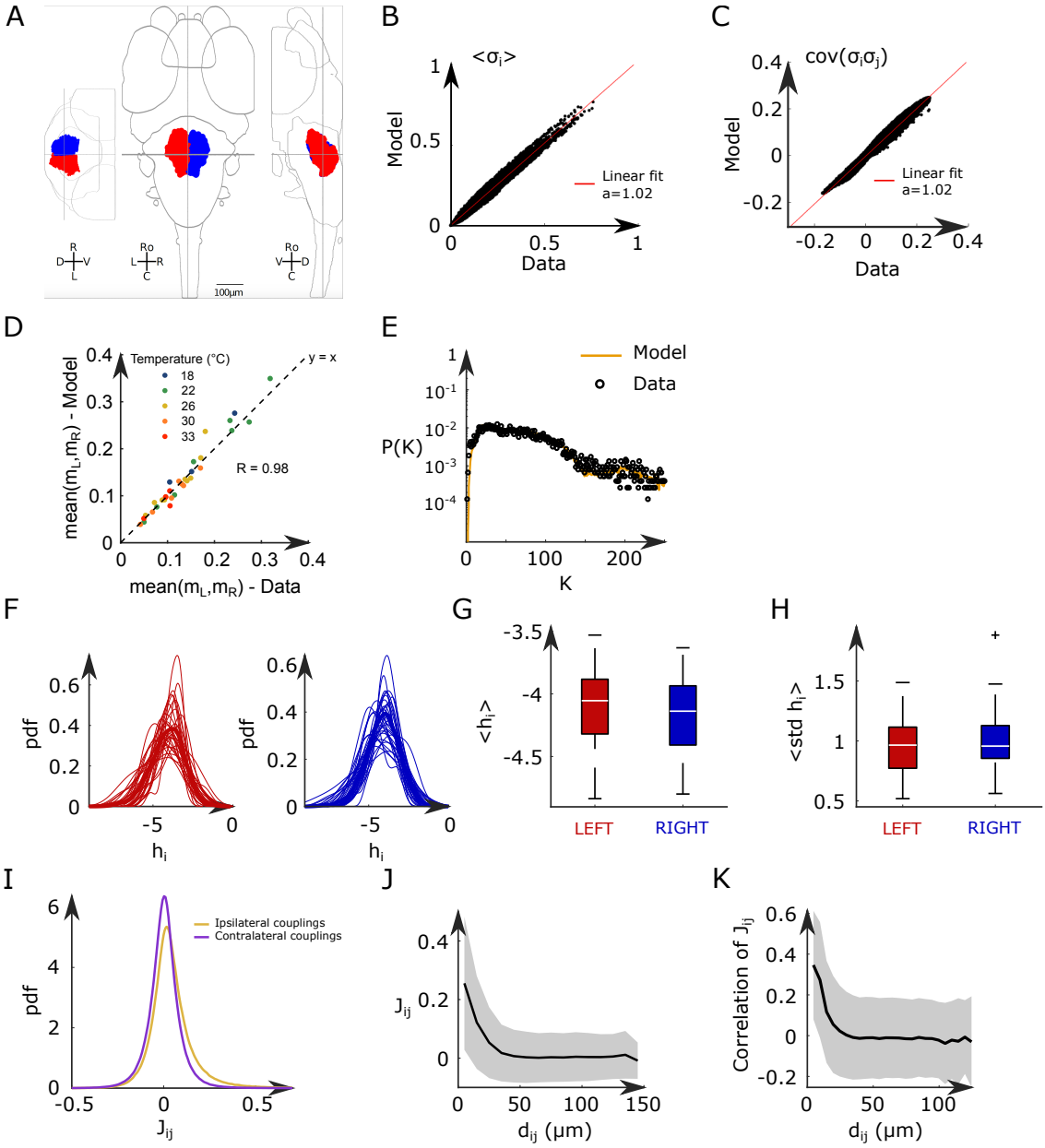

Figure S1: **Inference of the ARTR Ising model.**

A, Schematic view of a zebrafish larva brain. **B-C**, To assess the quality of the inference, we compare the mean activity (**B**) and the pairwise covariance (**C**) computed on the experimental and on the synthetic (model-generated) data (32 recordings, $n = 13$ fish). **D**, Mean activity of either side of the ARTR in simulations vs. experiments. Each dot refers to one fish at one temperature, colors encode temperature. **E**, Probability that K of the N neurons in the ARTR are active simultaneously in the data (black dots) and in the model (yellow line). To obtain these probabilities, we pool together all states where any K neurons are active together while the rest of the neurons are silent. **F**, Probability density function of the biases among the left (left panel) and right (right panel) subpopulations. Each thin line correspond to one experiment. The bold line corresponds to all the pooled data. **G**, Box plot across experiments of the average value of the biases for the left/right neurons of the ARTR. **H**, Box plot across animals of the standard deviation of the biases for the left/right neurons of the ARTR. **I**, Probability density function of the functional connectivity for the ipsilateral (purple line) and the contralateral (gold line) couplings. Plain line corresponds to the pdf across all animals. **J**, Functional connectivity as a function of the distance between neurons. **K**, Correlation between the couplings J_{ki} and J_{kj}, between one neuron k and two contralaterally located neurons i, j as a function of their distance d_{ij}.

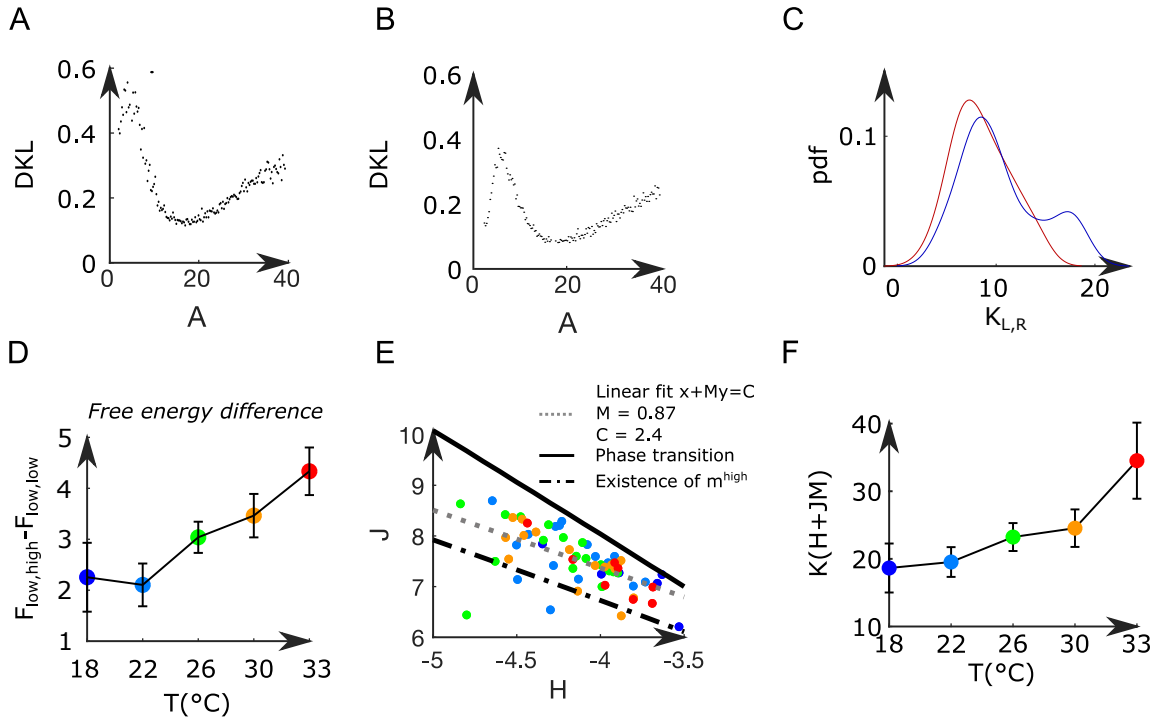

Figure S2: **Mean-field model of the ARTR.**

A-B, Kullback-Leibler divergence between the experimental and the Langevin distributions as a function of A (see Methods) for two data sets. **C**, Probability density function of K_R (blue line) and K_L (red line) across all recordings. **D**, Free-energy difference between stationary sates of the landscape as a function of the temperature. **E**, Phase diagram of the mean field model for $I = 0$ in the (H, J) plane. Each point corresponds to one experimental session and to one subregion (left or right). Same color code for the bath temperature as in the main figures. Below the dash-dotted line the unique stationary point of the free energy is m^{low}; above the m^{high} state is also present. The solid black line defines the first order transition of our model, where the free energies of the m^{low} and m^{high} states coincide. The dotted line corresponds to the best linear fit between J and H, with slope $-M = -0.87$. **F**, Average values (for all experiments and regions) of $K(H + J\,M)$ as a function of the temperature of the bath. Error bars are standard error of the mean.

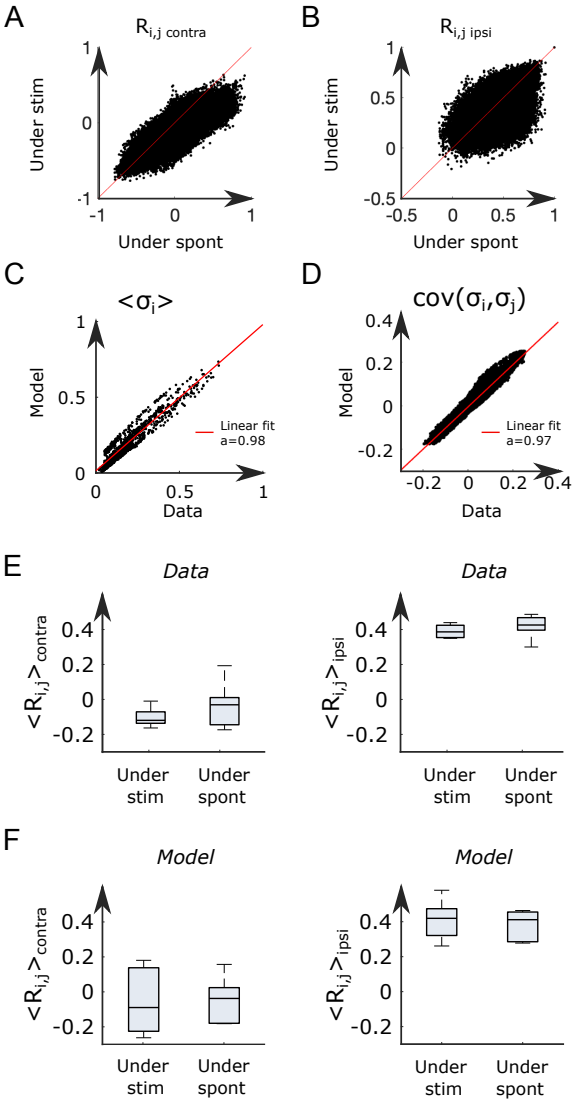

Figure S3: **A modified Ising model explains visually-driven properties of the ARTR.**
A, Scatter plot of the correlation between contralateral pairs of neurons under visual stimulation versus under spontaneous activity on $n = 6$ fish. **B**, Scatter plot of the correlation between ipsilateral pairs of neurons under visual stimulation versus under spontaneous activity. **C-D**, To assess the quality of the model of the ARTR of these visually-driven experiments, we compare the mean activity (**C**) and the pairwise covariance (**D**) computed on real data (spontaneous part of the recordings) to those computed on the synthetic data. **E**, Average Pearson correlation in the experimental recordings between contralateral and ipsilateral pairs of cells during stimulated or spontaneous activity ($n = 6$ fish). **F**, Average Pearson correlation in the simulated activity of the ARTR between contralateral and ipsilateral pairs of cells during stimulated or spontaneous activity ($n = 6$ fish).

Temperature (°C)	ID	Line	Age (dpf)	N_L	N_R	Acquisition rate (Hz)	Duration (s)
18	12	NucFast	6	146	180	5	1200
18	13	NucFast	7	37	96	8	1200
18	14	NucFast	6	179	174	8	1200
22	2	Nuc slow	7	177	212	3	1106
22	3	NucFast	5	152	85	3	1812
22	5	NucFast	5	158	123	5	1500
22	6	NucFast	5	98	134	5	1500
22	7	NucFast	6	122	221	5	1500
22	11	NucFast	6	295	320	5	1200
22	13	NucFast	7	37	96	8	1200
22	14	NucFast	6	179	174	8	1200
26	2	Nuc slow	7	177	212	3	1812
26	3	NucFast	5	152	85	3	1812
26	4	NucFast	5	110	76	3	1812
26	5	NucFast	5	158	123	5	1500
26	6	NucFast	5	98	134	5	1500
26	7	NucFast	6	122	221	5	1500
26	11	NucFast	6	295	320	5	1200
26	13	NucFast	7	37	96	8	1200
26	14	NucFast	6	179	174	8	1200
30	2	Nuc slow	7	177	212	3	1812
30	4	NucFast	5	110	76	3	1812
30	5	NucFast	5	158	123	5	1500
30	6	NucFast	5	98	134	5	1500
30	7	NucFast	6	122	221	5	1500
30	13	NucFast	7	37	96	8	1200
30	14	NucFast	6	179	174	8	1200
30	15	NucFast	7	202	252	8	1200
33	14	NucFast	6	179	174	8	1200
33	15	NucFast	7	202	252	8	1200
33	16	NucFast	6	127	123	7	1200
33	17	NucFast	5	62	170	10	1200

Table S1: Datasets properties.

Temperature (°C)	ID	J_L	J_R	I	H_L	H_R	K_L	K_R
18	12	7.06	7.23	-0.6	-3.66	-3.63	6.51	8.03
18	13	6.2	7.84	0.6	-3.53	-4.34	3.18	8.27
18	14	7.27	7.24	0.31	-3.88	-3.99	11.04	10.74
22	2	8.2	8.28	0.12	-4.24	-4.23	6.65	7.96
22	3	8.18	7.14	0.55	-4.26	-4.13	9.38	5.24
22	5	7.59	7.01	0.4	-4.03	-3.8	5.56	4.33
22	6	7.13	8.69	1.1	-4.49	-4.64	5.21	7.12
22	7	7.09	7.46	0.43	-3.73	-3.95	6.28	11.39
22	11	7.82	7.59	-0.1	-4.07	-3.91	8.28	8.98
22	13	6.54	7.82	1.45	-4.29	-4.5	7.11	18.46
22	14	7.41	8.03	0.47	-4.28	-4.43	10.91	10.6
26	2	8.37	8.22	-0.49	-4.47	-4.31	9.72	11.64
26	3	8.42	7.49	0.53	-4.56	-4.62	8.26	4.61
26	4	8.63	6.44	0.85	-4.83	-4.79	10.37	7.16
26	5	7.29	7.59	0.48	-3.92	-4.14	9.08	7.06
26	6	7.43	7.86	0.41	-3.99	-4.1	8.59	11.75
26	7	7.55	7.96	0.32	-4.08	-4.22	4.45	8.06
26	11	7.27	7.45	0.37	-3.89	-3.92	10.31	11.18
26	13	6.99	7.3	0.6	-3.99	-3.94	6.37	16.55
26	14	7.91	7.35	0.5	-4.34	-4.16	11.32	11.01
30	2	7.54	7.96	-0.12	-4.54	-4.56	7.02	8.41
30	4	8.36	7.73	0.11	-4.52	-4.18	9.64	6.66
30	5	6.77	6.42	0.66	-3.8	-3.87	9.18	7.15
30	6	7.35	7.38	0.45	-3.91	-3.97	7.53	10.3
30	7	7.43	8.07	0.42	-3.93	-4.38	7.09	12.84
30	13	6.91	7.41	0.73	-4.13	-4.03	5.78	15
30	14	7.51	7.45	0.11	-3.87	-3.89	9.42	9.15
30	15	8.01	8.33	0.58	-4.45	-4.46	13.83	17.26
33	14	6.74	7.02	0.76	-3.8	-3.97	9.32	9.06
33	15	6.99	7.47	-0.02	-3.68	-3.91	14.85	18.52
33	16	7.53	8.25	-0.11	-4.16	-4.43	14.43	13.97
33	17	6.66	7.36	0.45	-3.69	-3.89	11.92	32.69

Table S2: Parameters of mean-field models.

ID	J_L	J_R	I	H_L	H_R	K_L	K_R
1	7,54	7,35	-0,67	-3,75	-3,44	5,60	3,43
2	7,10	7,42	0,64	-3,69	-4,02	7,91	12,82
3	7,51	7,92	-0,28	-3,96	-4,08	4,98	3,90
4	8,38	6,25	-0,04	-3,68	-3,18	13,33	4,44
5	8,73	8,24	0,01	-4,38	-4,13	6,11	6,89
6	7,87	7,71	0,51	-4,17	-4,09	16,19	15,52

Table S3: Parameters of the mean-field model for two-photon light-sheet data sets from (Wolf et al., 2017)