Spiking attractor model of motor cortex explains modulation of neural and behavioral variability by prior target information

Vahid Rostami*[1], Thomas Rost*[1], Felix J. Schmitt[1], Sacha J. van Albada[3,4,5], Alexa Riehle[2,3], Martin P. Nawrot[1]

1 Computational Systems Neuroscience, Institute of Zoology, University of Cologne, Cologne, Germany

2 UMR7289 Institut de Neurosciences de la Timone (INT), Centre National de la Recherche Scientifique (CNRS)—Aix-Marseille Université (AMU), Marseille, France

3 Institute for Advanced Simulation (IAS-6), Jülich Research Center, Jülich, Germany

4 JARA Institute Brain Structure-Function Relationships (INM-10), Jülich Research Center, Jülich, Germany

5 Computational Neuroanatomy, Institute of Zoology, University of Cologne, Cologne, Germany

Correspondence: martin.nawrot@uni-koeln.de

Abstract

When preparing a movement, we often rely on partial or incomplete information, which can decrement task performance. In behaving monkeys we show that the degree of cued target information is reflected in both, neural variability in motor cortex and behavioral reaction times. We study the underlying mechanisms in a spiking motor-cortical attractor model. By introducing a novel and biologically realistic network topology where excitatory neuron clusters are locally balanced with inhibitory neuron clusters we robustly achieve metastable network activity across a wide range of network parameters. In application to the monkey task, the model performs target-specific action selection and accurately reproduces the task-epoch dependent reduction of trial-to-trial variability *in vivo* where the degree of reduction directly reflects the amount of processed target information, while spiking irregularity remained constant throughout the task. In the context of incomplete cue information, the increased target selection time of the model can explain increased behavioral reaction times . We conclude that context-dependent neural and behavioral variability is a signum of attractor computation in the motor cortex.

1 Introduction

Despite correct motor performance in well-trained animals, behavioral reaction times and neural spiking activity in the motor cortex appear highly variable across repetitions of an identical task[1]. Hence, any successful biologically realistic model of cortical motor control must provide a mechanistic explanation for both types of variability.

Over the past decades, attractor dynamics has been established as the most viable mathematical concept and computational model to support working memory and decision making in sensory-motor tasks, and there is considerable experimental evidence to support it[2–10]. Yet, despite the fact that the basic theory of attractor dynamics is well understood[11–13], its robust implementation and functional application in biologically realistic spiking neural network models remains a challenge.

The classical cortical model is the balanced network of excitatory and inhibitory neurons with random connectivity[14,15]. This model successfully captures certain aspects of cortical spiking dynamics, including low spontaneous firing rates and irregular spiking statistics. However, it fails to explain the observed cortical trial-to-trial variability and cannot accommodate metastable attractors. Metastable attractors (or simply metastability) refers to the phenomenon where the neural dynamics exhibit transient states of coordinated activity across ensembles of neurons, known as metastable states. These states persist for durations ranging from tens to hundreds of milliseconds before transitioning to the next state. The presence of such metastable sequences is widespread across various cortical areas and behavioral contexts, prompting investigations into their functional role in sensory and cognitive processing[9,16–20]. Each metastable state is characterized by a specific pattern of ensemble activity, involving the firing rates of a collection of neurons. The transitions between these states are marked by abrupt, jump-like modulations, presenting neural activity as a sequence of discrete and quasi-stationary states. This phenomenon is observed in response to external stimuli as well as during spontaneous activity.

Over recent years, the balanced random network has been extended to accommodate

strongly interconnected neuron clusters. With carefully tuned network parameters, this architecture introduces metastable attractor dynamics and can capture aspects of cortical trial-to-trial variability dynamics[16,17,21–23]. These studies used purely excitatory neuron clustering but neglected any structure in the topology of local inhibitory networks.

Despite the vital role of inhibitory neurons in cortical dynamics, the circuit connectivity of inhibitory neurons has remained poorly understood until recent years. Several influential studies provided evidence that inhibitory interneurons connect non-specifically to surrounding excitatory pyramidal cells[24,25], which inspired the term 'blanket of inhibition' and supported the rationale behind a purely excitatory cluster topology[21,22,26]. However, additional studies now provide a more complete picture, suggesting a high degree of specificity and possible clustering of inhibitory neurons[27–35]. In particular, it has been argued, based on anatomical and physiological evidence, that inhibitory networks can be strongly interconnected locally. Moreover, neurons that receive strong excitatory input typically also receive strong inhibitory input, supporting local balancing at the level of single cell input[27,36,37]. In addition, recent theoretical studies have corroborated the importance of inhibition in attractor type dynamics and found, based on the analytical treatment of binary and firing rate models, that inhibitory clustering strongly improves the robustness of metastable attractor dynamics[34,38].

In the present study, we propose a novel network architecture for spiking cortical attractor networks using combined excitatory and inhibitory clustering. We show that inhibitory clustering maintains the local balance of excitation and inhibition and yields the desired metastability robustly over a wide range of network sizes and parameters. We utilize our model to mechanistically explain task-related dynamics of multiple single-unit activity recorded from the motor cortices in two monkeys (M1 and M2) during a delayed reaching task. Here we consider the decision-making process and do not attempt to generate muscle-like signals to generate movement trajectories. We find that our model qualitatively and quantitatively captures *in vivo* firing rates, task epoch related dynamics

of trial-to-trial variability and spiking irregularity, and behavioral reaction times. Variation of the behavioral task involved different levels of target uncertainty during the delay period and resulted in corresponding levels of neural trial-to-trial variability and the systematic variation in behavioral reaction times, both in the monkey data and in our model simulations.

2 Results

Spiking networks with local excitatory-inhibitory clusters can explain cortical variability dynamics in vivo. We start out with analyzing the temporal dynamics of spike train variability in single-unit recordings from the monkey motor cortex during a delayed center-out reach task[39,40] (see Methods). The monkeys were instructed to reach for one of six target buttons at the end of an instructed delay period during which a varying degree of target certainty was cued by either one, two or three adjacent targets. We first consider the simplest task condition in which complete target information was provided to the monkey with the onset of the preparatory period (preparatory signal, PS, indicated by the green circle in Fig. 1 a). After a fixed delay of one second the monkey was prompted to execute the movement by the response signal (RS, red circle). Correct execution of a trial was rewarded with fruit juice.

Common statistical measures for quantifying the variability dynamics of spiking neurons are the Fano Factor (FF) and the local coefficient of variation (CV_2) of inter-spike intervals[41–43]. The FF determines the neural response variability across the repetition of trials while the CV_2 measures the variability of the inter-spike intervals and thus quantifies the irregularity of spike occurrences within each trial. On theoretical ground, the expectation values for both measures are interdependent (see Methods) and their empirically estimated absolute values and relation can be interpreted with respect to the nature of the underlying generative neuronal and network processes (see Discussion).

We apply FF and CV_2 in a time-resolved fashion using a sliding observation window[42] (see Methods). Fig. 1 a shows that during spontaneous activity and before cue onset (PS) trial-to-trial variability is high (FF ≈ 1.8). After PS, the FF decreases significantly (red curve in Fig. 1 a) before reaching a constant plateau. This task-related reduction of the FF has been demonstrated previously in motor cortices and for different behavioral tasks[40,44–46] as well as in other cortical areas[45,47]. The irregularity of inter-spike intervals, CV_2, on the other hand, remains constant over time (black curve in Fig. 1 a) and does not show any dependency on the experimental epochs. For the Poisson process the expectation is FF $= CV_2 = 1$ (dashed gray line in Fig. 1 a).

Next, we study variability dynamics in a biologically plausible spiking network model of the cortex with excitatory cluster topology. Deviating from the random balanced network, the excitatory neuron population is divided into subpopulations with strong internal connectivity while excitatory connections between clusters are comparatively weak[16,21–23]. The model is composed of 4.000 excitatory (E) and 1.000 inhibitory (I) exponential integrate-and-fire neurons. The E neurons are organized into $Q = 50$ recurrent clusters (Fig. 1 b top panel). Synaptic connections within each excitatory cluster are potentiated by a factor J_{E+}, and, to maintain overall balance, connections between E neurons belonging to different clusters are depressed by a factor J_{E-} (see Methods). Inhibitory neurons are non-specifically connected to themselves and to E neurons (see Methods). A consistent reduction in FF upon stimulation can only be achieved with a strong stimulus (dark red curve) whereas we observe an inconsistent behavior for intermediate stimuli or even an increase in FF for stimuli of low strength (light red curve), conflicting with our experimental observation. As for the dynamics of irregularity, we observe a reduction in CV_2 during stimulation that tends to be stronger for a stronger stimulation. This stimulus-induced increase in regularity is again clearly inconsistent with our experimental observation (Fig. 1 a) and, to our knowledge, has not been reported in any other experimental study. In summary, the E-clustered network is inconsistent with the experimental observations in

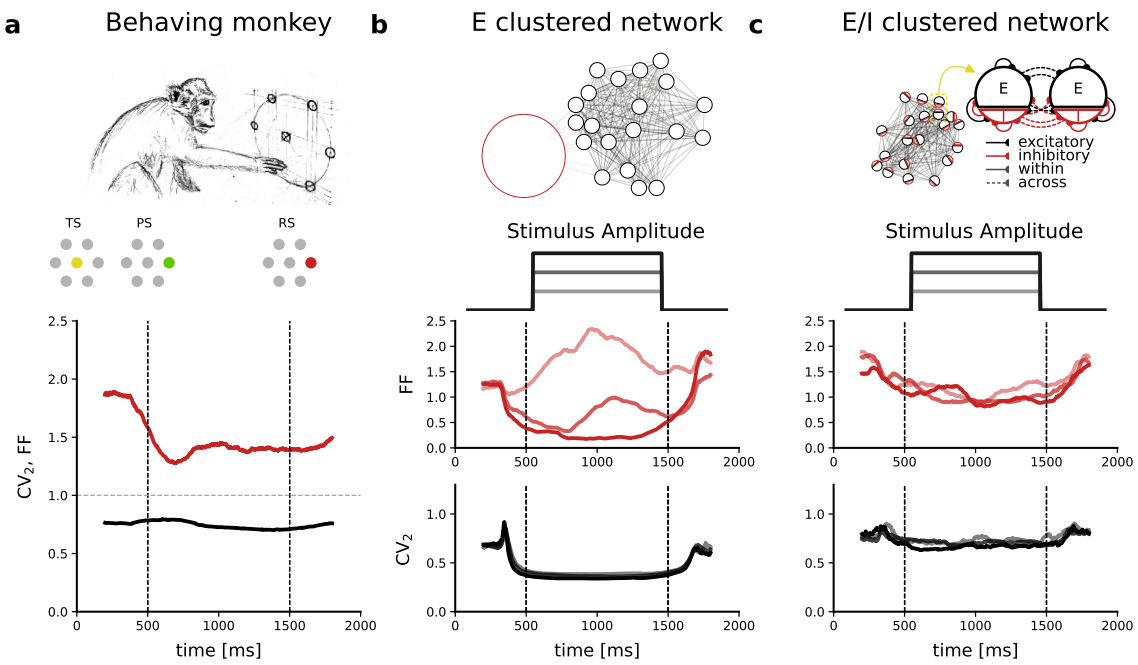

Figure 1: **Variability dynamics observed experimentally and in clustered spiking network models. a)** Experimental data recorded from the motor cortex of M1 during a delayed center-out reach task[40]. After the monkey places its hand on the central touch-sensitive LED the trial start (TS) is indicated by lighting up the LED (yellow). At t=500 ms the preparatory stimulus (PS) indicates the movement target (target LED is lit in green). After a fixed preparatory delay period of 1.000 ms, during which the monkey is not yet allowed to move his hand, the response stimulus (RS) is indicated by changing the color of the target LED to red and the monkey is allowed to move. The time-resolved FF and CV_2 averaged across neurons are shown in red and black, respectively. Gray horizontal dashed line shows FF and CV_2 for the Poisson process. . **b)** Network model with purely excitatory clusters and global inhibition (E-clustered network). Clusters in the network receive step-wise external input with three different stimulus amplitudes (0.2 pA, 0.25 pA, 0.3 pA) as indicated in shades of gray. The corresponding time-resolved FF is shown in shades of red and the time-resolved CV_2 in shades of gray. **c)** Same analyses as in (b) but for the proposed network model with excitatory and inhibitory clusters (E/I-clustered network). FF and CV_2 were computed in a 400 ms sliding window; Note that we use a centered window and therefore the curve can start decreasing before the indicated trigger (max 200 ms); simulated data comprised 50 trials.

two ways: (1) a reduction in count variability (FF) is achieved only with a strong stimulus, while a weak stimulus can lead to an increase in FF; (2) during network stimulation the irregular spiking is disrupted, and the CV_2 assumes unrealistically low values.

To match the experimentally observed stimulus-induced variability dynamics, we suggest a novel type of network connectivity. Recent anatomical and physiological studies point to a high local connectivity and a possible clustering of inhibitory neurons[27–29,31–34]. We therefore combine excitatory and inhibitory clustering in our spiking network model following our previous proposal for binary networks[38] (Fig. 1 c top panel). We simulate this E/I-clustered network model using the same parameters as for the E-clustered network (see Methods). Fig. 1 c shows that the trial-to-trial variability (FF) decreases robustly even for a weak network stimulation while the CV_2 does not show any stimulus-induced changes, matching our experimental observations (Fig. 1 a). We will next investigate the network mechanisms that support realistic variability dynamics in E/I-clustered but not in E-clustered networks.

Metastability emerges robustly in the E/I-clustered spiking network. We first ask which network parameters determine the emergence of metastability and winnerless competition among embedded clusters. To answer this question we examined the effect of two important clustering parameters, the clustering strength J_{E+} and the number of clusters Q for a fixed network size. Metastability is expressed in the successive activation and inactivation of individual neuron clusters where individual neurons switch between lower and higher firing rate states. Across repeated observations of individual neurons we therefore expect a high variation of the spike count. Thus, the FF provides a proxy measure for metastability.

We again consider a network of 4.000 excitatory and 1.000 inhibitory neurons (Table 1). We first kept the number of clusters $Q = 50$ fixed and varied the cluster strength between $J_{E+} = 1$, which coincides with the classical random balanced network without clustering, to

$J_{E+} = Q = 50$ with zero excitatory coupling between different clusters (see Methods). The E-clustered network (Fig. 2 a) can show metastability as indicated by the alternating firing rate states of individual cluster populations (insets in Fig. 2 a) only in a very narrow range of excitatory cluster strengths around $J_{E+} \approx 3.5$, which is accompanied by correspondingly large Fano factors (FF > 1). When increasing J_{E+}, the cluster dynamics rapidly breaks down and the FF falls below its initial value associated with the random balanced network ($J_{E+} = 1$) and eventually the network gets stuck in a single state with a few clusters becoming permanently active with high firing rates and regular spiking patterns.

A different picture emerges in the E/I-clustered network (Fig. 2 c). Metastability is achieved over a wide range of J_{E+}. With increasing cluster strength, the duration of individual cluster activations becomes longer. For large values $J_{E+} \gtrsim 15$ the variance over repeated simulations is high, as some cases exhibit extensive cycling between clusters while in other cases one or a few clusters become dominantly active and suppress the winnerless competition dynamics. At $J_{E+} = Q = 50$ coupling between the populations exists only through strong inhibitory connections and most populations are silenced by a few winning clusters. The desired attractor dynamics thus takes place at lower values of J_{E+}, which yield realistic average trial-to-trial variability in the approximate range $1 <$ FF < 3.

In a next step we in addition varied the number of clusters Q while keeping the total number of neurons N fixed. This changes the size of each cluster, i.e. larger Q means a smaller number of neurons per cluster. For the purely excitatory cluster topology (Fig. 2 b) we find that there is only a very small parameter region that shows the desired high spike count variability across repeated observations (FF > 1). Thus, the E-clustered network is able to facilitate winnerless competition, but stable switching dynamics can be achieved only by extensive parameter tuning. Litwin-Kumar and colleagues[21] suggested a variation of their E-clustered network architecture where each excitatory cluster preferentially excites one disjunct pool of neurons within the total inhibitory population. Repeating our calibration for this alternative architecture, we again obtained

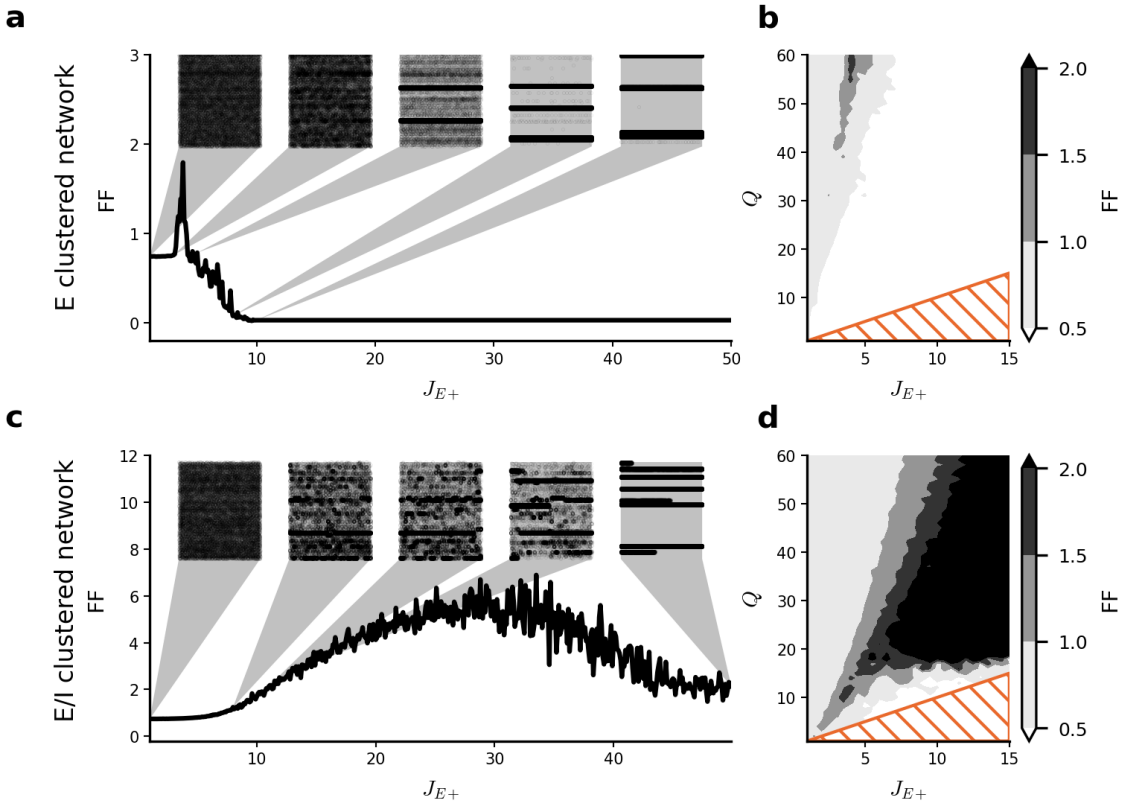

Figure 2: **Excitatory-inhibitory clustering facilitates winnerless competition across a wide range of network parameters. a, c)** FF versus J_{E+} for networks with $Q = 50$ clusters computed from 20 trials of 400 ms duration and averaged over 50 network realizations. Insets show two seconds of single trial network activity for all excitatory neurons grouped by their cluster identity. **b, d)** Effect of cluster strength J_{E+} and number of clusters Q on metastability and FF. Shaded orange triangles indicate the zone below $J_{E+} = Q$, where the clusters are completely decoupled. Panels **a, b** show results for the E-clustered network (see also Supplemental Fig. S1), panels **c, d** for the E/I-clustered network. Parameters apart from J_{E+} and Q as in Table 1. Insets in **a** for $J_{E+} = (1, 3, 3.7, 8, 10)$, in **c** for $J_{E+} = (1, 8, 10.5, 14, 50)$.

the same result (Supplemental Fig. S1) as for the E-clustered network (Fig. 2 b). In contrast, in the E/I-clustered network attractor dynamics as indicated by a high FF > 1 emerges robustly over a large range of cluster numbers Q and excitatory cluster strengths J_{E+} as shown in Fig. 2 d. This means that metastability in the E/I-clustered network is not sensitive to variations or perturbations in network parameters. Importantly, the E/I-clustered topology supports metastability even for large clusters in contrast to the E-clustered network where attractor dynamics breaks down for larger cluster sizes (lower values of Q).

Local balance of excitation and inhibition facilitates attractor dynamics and maintains spiking irregularity. Synaptic excitation and inhibition are opposing effects that together determine the activity and maintain the excitability of cortical neurons and networks[36,48,49]. A number of physiological and model studies have shown that excitatory and inhibitory synaptic inputs retain a fixed and roughly equal proportion during both spontaneous and evoked states[37,50–54]. This mechanism known as balance of excitation and inhibition is crucial for maintaining irregular spiking and stable network activity, and disruption of this balance has been linked to pathological activity states such as epileptic seizures[53].

In this section, we study the balance of excitation and inhibition in E- and E/I-clustered networks. Using again the parameters in Table 1 we simulate 1.000 ms of spontaneous activity of E- and E/I-clustered networks with $Q = 50$ clusters. Fig. 3 illustrates the difference between E-clustered (left panels) and E/I-clustered (right panels) networks on the level of individual neurons. The raster plot in Fig. 3 a shows the activity of nine sample excitatory clusters in the E-clustered network. Neurons across all nine clusters exhibit low firing rates until at about t = 350 ms one cluster switches into an activated state. The corresponding instantaneous firing rate averaged across all neurons in the activated cluster increases strongly to almost 90 spikes/s, whereas the firing rates of the other clusters

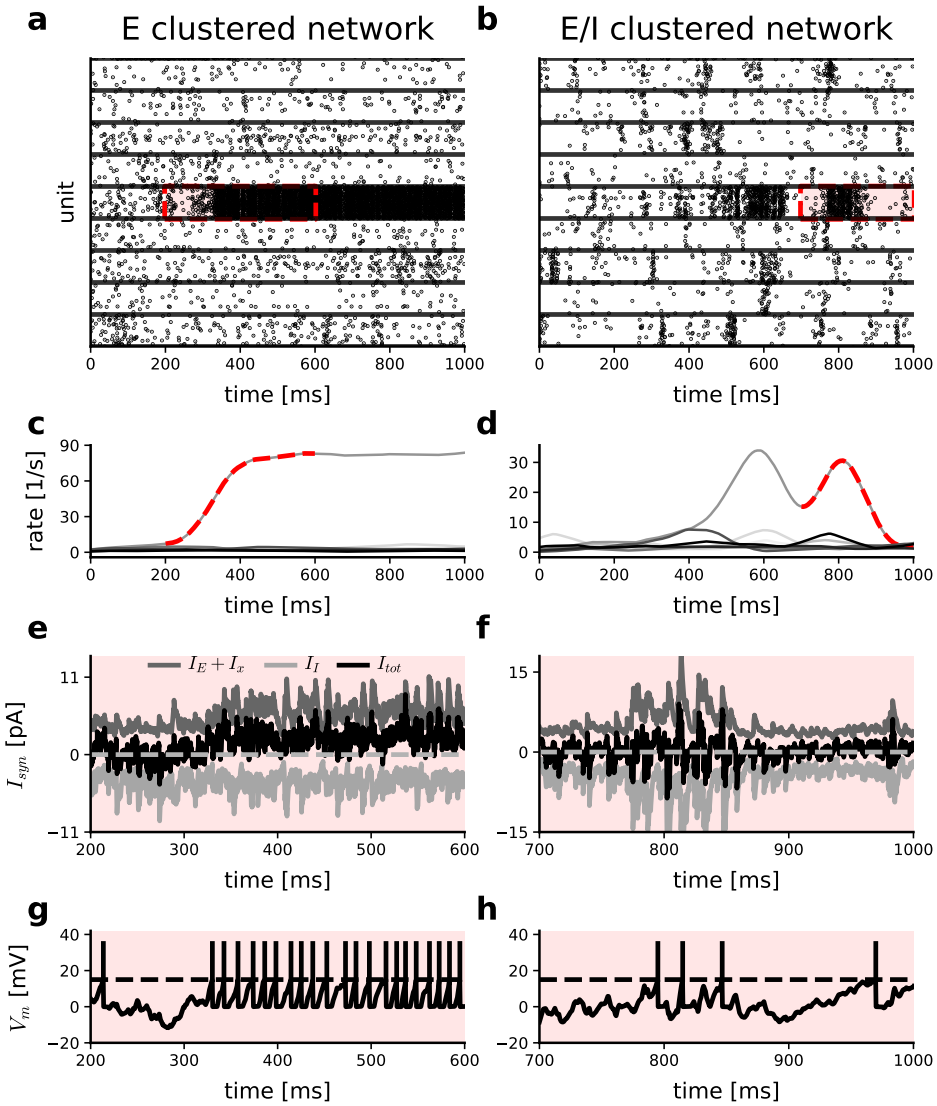

Figure 3: **E/I-clustering maintains local balance of excitation and inhibition and irregular spiking.** Comparison of E-clustered (left) and E/I-clustered (right) networks. **a, b)** Spike raster display of nine excitatory clusters during 1.000 ms of spontaneous activity. Horizontal lines separate different clusters. Red shaded area indicates the epoch of interest where a switch into an active state occurred in one cluster. **c, d)** Average firing rate of each of the nine clusters shown above as estimated with a 50 ms triangular kernel[55]. Dashed lines correspond to the red shaded region of interest in the upper panels. **e, f)** Synaptic currents in a randomly chosen unit around the switch to the active state within the cluster and epoch of interest as indicated above. **g, h)** Membrane potential for the unit shown in **e** in the same time interval.

remain low. Fig. 3 e shows the synaptic input currents to one single neuron in the activated cluster around the switching time (red shaded interval in Fig. 3 a). While the excitatory input current increases (I_E, upper trace) due to the strong mutual excitation within this cluster, the inhibitory input (I_I, lower trace) remains constant. As a result, the net input current (I_{tot}, middle trace) increases to positive values and hence the neuron operates in the mean-driven rather than in the fluctuation-driven regime. As a result, a large portion of the neurons in the active cluster fire with high rates and high regularity, as can be seen in the membrane potential of the example neuron depicted in Fig. 3 g. This provides the mechanistic explanation for our results of a strongly decreased CV_2 in Fig. 1 b.

The right-hand side of Fig. 3 analyzes the equivalent scenario for an E/I-clustered network. The raster plot in Fig. 3 b indicates that in this spontaneous state individual clusters can switch from an inactive state with low ongoing firing rates to activated states of variable duration. This attractor dynamics involves moderate firing rates during activated states (Fig. 3 d), which are considerably lower than the high firing rates observed in the E-clustered network Fig. 3 d. The synaptic input currents in a sample neuron are shown in Fig. 3 f, again at the transition from the inactivated to the activated state of its cluster. In effect, inhibitory clustering ensures that the inhibitory synaptic currents increase in parallel with the excitatory synaptic currents during cluster activation, thereby increasing the variance but not the mean of the net synaptic input, which is kept balanced throughout. Hence, all neurons remain in the fluctuation-driven regime and retain irregular spiking as illustrated in Fig. 3 h and shown for the network population in Fig. 1 c. In conclusion, maintaining balance of excitatory and inhibitory inputs to lower firing rates in activated clusters supports switching between inactive and active states by decreasing the rate differences of both, facilitating the desired metastable dynamics. In purely excitatory clusters the unbalanced input together with the strong connectivity among neurons within a cluster results in high firing rates of the excitatory neurons that operate in a mean-driven input regime. The increased self-excitation within an excitatory cluster and the lack of

balancing inhibition leads to longer duration of an activated cluster state and can result in permanent winner-takes-all scenario where a single cluster remains activated for a very long duration (Fig. 3 a).

E/I-clustering increases the dynamic range of the stimulus-response function. We have seen that the E-clustered network fails to capture the reduction in trial-to-trial variability during stimulation with a weak stimulus (Fig. 1). Here we ask what exactly constitutes a weak stimulus by analyzing in detail how the cluster response rate and the change in trial-to-trial variability depend on the stimulus amplitude in networks with and without inhibitory clustering.

To this end we stimulated five out of 50 clusters by means of a constant input current injected into all neurons belonging to these clusters (Fig. 4 a). In our analysis we compare network activity in the spontaneous state (before stimulation) with the network activity during the evoked state (during stimulation) and calculate the changes in firing rate (Δrate) and in trial-to-trial variability (ΔFF) as a function of stimulus strength. In Fig. 4 b we show the relation of stimulus amplitude and average response of the activated clusters separately for the neurons in the stimulated and in the non-stimulated clusters. In the stimulated clusters of the E/I network we observe a smooth relation between input current and firing rate increase across the tested stimulus range. In contrast, for the case of a purely excitatory cluster topology the firing rate in stimulated clusters increases steeply within a narrow range of stimulus current amplitude ($\approx 0.05 - 0.08\,\mathrm{pA}$) and continues to increase rapidly with increasing input current. The non-stimulated clusters exhibit a small decrease in their average firing rate during stimulation.

The dynamics of trial-to-trial variability again shows a clear difference in the two different network topologies. In the E-clustered network, there is a stark increase in FF for stimulus amplitudes in the lower half of the stimulus range of up to $\sim 0.4\,\mathrm{pA}$ (Fig. 4 c). The desired effect of a reduced FF is achieved only for increasingly large input

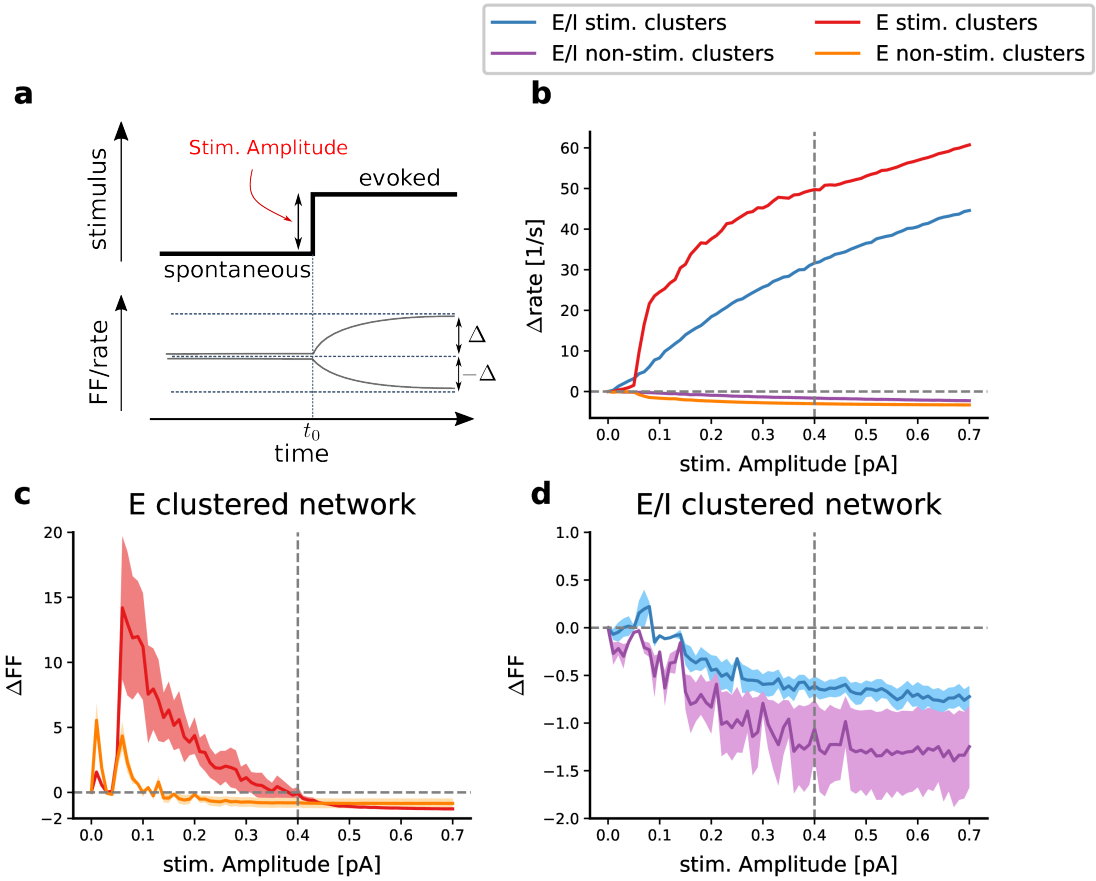

Figure 4: **Stimulus-evoked changes in firing rate and trial-to-trial variability. a)** Cartoon depicting the stimulus-induced transition from the spontaneous to the evoked cluster state. Upper panel shows the step change in the external input current at time t_0. Lower panel depicts the stimulus-induced changes in the average firing rate or FF relative to the spontaneous level. Stimulation is applied to 5 out of the total of $Q = 50$ clusters, leaving 45 non-stimulated clusters. **b)** Average stimulus-response function Δrate versus the amplitude of the injected step current. Only periods of time within the average cluster rate that exceed the median rate of the non-stimulated clusters are included. Different colors indicate the behaviors for stimulated and non-stimulated clusters in the E and E/I-clustered networks as indicated in the legend. **c,d)** ΔFF versus stimulus amplitude for the E-clustered network (c) and the E/I-clustered network (d) estimated across 50 trials. Both measures are calculated for 1000 ms of spontaneous and evoked activity, respectively. The difference is computed for each neuron separately before averaging across neurons. Firing rate differences of single units were averaged across the 50 trial repetitions. Shaded areas show the standard error of the mean calculated across the stimulated and non-stimulated clusters, respectively. Network parameters as in Table 1.

currents that correspond with high cluster rate responses of Δrate$>$ 40 spikes per second (cf. vertical dashed line in Fig. 4 b and c). This means that the E-clustered network fails to reproduce the experimentally observed reduction in trial-to-trial variability for weak stimuli and moderate average neuronal response rates. The reason for the increase in variability for weak to moderate stimulus amplitudes is that stimulated clusters switch into an activated state in some trials but fail to do so in others, i.e. cluster activation is unreliable. This coincides with our theoretical prediction based on the analytic treatment and numerical simulation of binary cluster networks where stable fixed points require considerable excitatory input stimulation and can assume only high cluster rates[38].

In contrast, the E/I topology shows a reduction in trial-to-trial variability of single-neuron spiking in stimulated and non-stimulated clusters even for weak stimuli and correspondingly low average cluster response rates (Fig. 4 b). This reflects that cluster stimulation reliably initiates switching into an excited state in each single trial. The range of ΔFF in the models closely matches the average ΔFF observed in the experimental data (cf. Fig. 1 a). We will show that, for matching the neuronal dynamics observed in the experimental *in vivo* data, weak stimuli are required with stimulus amplitudes that lie in the very low range of the stimulus amplitudes considered in Fig. 4.

In summary, the E/I-clustered network robustly captures the reduction in trial-to-trial variability during stimulation due to the reliable switching to the activated state even for low stimulus amplitude retaining a smooth relationship between stimulus strength and firing rate response over the complete stimulus range, while the E-clustered network fails to represent low to moderate population rates in activated clusters and cannot reproduce the reduction in trial-to-trial variability .

Motor cortical activity reflects target uncertainty during movement preparation and execution. We now analyze monkey behavior and motor cortical single-unit activity during the delayed reach task in its full complexity for monkeys M1 (Fig. 5) and M2

(Fig. S3). As shown in Fig. 5 a the monkeys performed the task under three different conditions that varied in the amount of cued target information. At the beginning of the 1 s delay period and in Condition 1, the PS cued a single target by a green light (full target certainty). The target light turned red at the end of the delay period (RS), prompting the monkey to move. In Condition 2, the PS cued two possible targets by two adjacent green lights, one of which was then randomly chosen as the final movement target (single red light) presented as RS. Condition 3 implied the highest target uncertainty during the preparatory period with three possible adjacent targets cued by the PS.

We first analyzed the encoding of movement direction in the motor cortical single-unit activity. To this end, and reconsidering the approach taken by Rickert et al.[40], we trained and cross-validated a classifier to predict the direction of the executed movement in each trial based on the neuronal population activity (see Methods). We then computed the decoding accuracy, i.e. the fraction of correctly predicted single-trial movement directions, as a function of trial time as shown in Fig. 5 b for M1. In all three task conditions and during the preparatory period the decoding accuracy reaches the theoretical limit that reflects the target information available to the monkey. When full information was available (Condition 1), decoding accuracy approaches unity. When target information was incomplete with either two or three possible targets indicated at PS, the decoding accuracy levels at 1/2 and 1/3, respectively, accurately reflecting target uncertainty. After the RS resolved target ambiguity by indicating the final single movement target, decoding accuracy approached unity in all three conditions.

Next, we asked whether task uncertainty modulates trial-to-trial variability of the single-unit activity. We therefore computed ΔFF as a function of trial time comparing the result across the three different task conditions for monkey M1 in Fig. 5 c (cf. Supplementary material for monkey M2). In all three conditions, spike count variability is initially quenched in reaction to the PS. Subsequently, the FF recovers to a different level depending on the experimental condition.

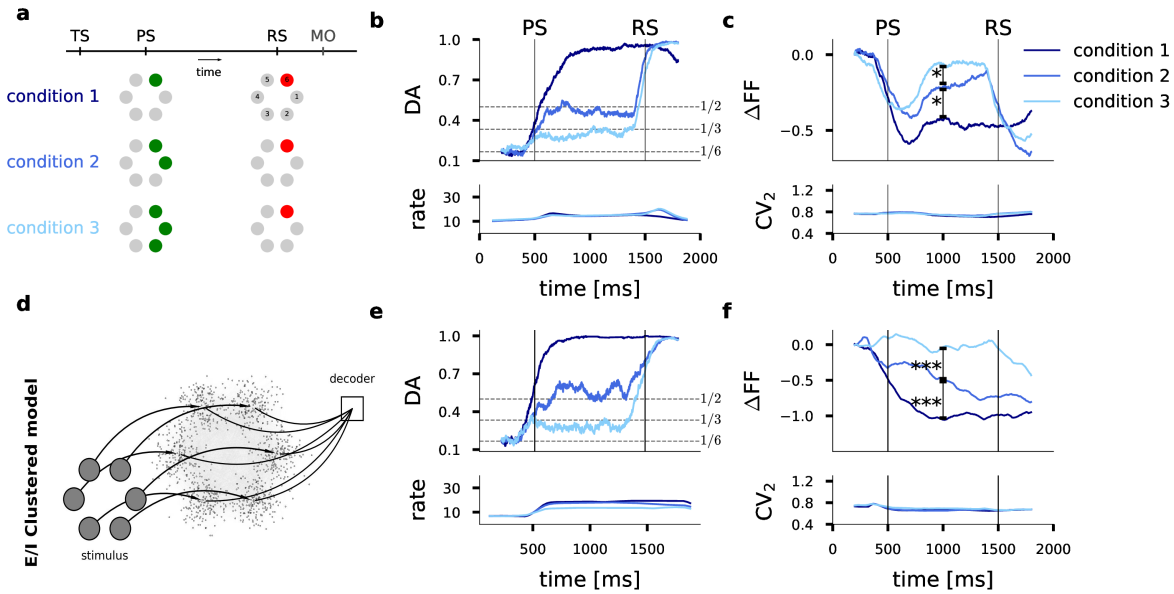

Figure 5: **Functional E/I-clustered network model captures context dependency of information encoding, firing rates and variability dynamics. a)** Full experimental protocol of the delayed center-out reach task with three experimental conditions. The preparatory signal (PS) indicated either one, two or three adjacent targets in green, corresponding to Condition 1 (full target information), and Conditions 2 and 3 (incomplete target information). After a fixed delay period of 1.000 ms the reaction signal (RS) resolved target ambiguity by indicating the final movement target in red, prompting the monkey to execute the movement. The movement onset time (MO) and movement end time (ME) are recorded in each single trial. **b)** Upper panel: Accuracy of decoding movement direction from the neuronal population activity for the three task conditions. Lower panel: population firing rates. **c)** Upper panel: Task-related reduction in trial-to-trial variability ΔFF as a function of trial time; $*$ indicates pairwise significant differences across FF distributions (Wilcoxon signed rank test, Conditions 1-2: $p = 0.007$, Conditions 2-3: $p = 0.03$). Lower panel: Time-resolved estimate of the CV_2. **d)** Architecture of the E/I-clustered attractor network model. Each of the six embedded clusters represents one target direction and can receive excitatory input with the PS and RS stimuli. Each cluster is assigned a direction for which it receives its external stimulus . To match the experimental firing rates in **b** we determined 0.1 pA as stimulus amplitude that was kept fixed in all three conditions. The decoder integrates the average cluster firing rates and generates a decision. **e, f)** Analysis as in b-c but for neurons in the spiking network model. Distributions of FF are pairwise significantly different as determined by a Wilcoxon signed rank test (Conditions 1-2: $p = 3.6 \times 10^{-13}$, Conditions 2-3: $p = 3.7 \times 10^{-30}$). Decoding accuracy, FF and CV_2 were estimated in a sliding window of 400 ms width.

In the case of complete target information (Condition 1) the FF remains at a low value, while in Condition 2 it resumes a higher, and in Condition 3 a still higher plateau value that still remains slightly below the baseline level during spontaneous activity. The differences between the task conditions are statistically significant. Thus, the level of trial-to-trial variability of single-neuron spiking activity during the preparatory period directly corresponds to the degree of target ambiguity, with higher uncertainty leading to increased variability and lower uncertainty resulting in reduced variability.. Following the RS and during movement execution, the FF shows an additional reduction in Conditions 2 and 3. No additional reduction in FF is visible in the single-target condition. Note that the average single-neuron spiking irregularity (CV_2) in Fig. 5 c (lower panel) remained constant throughout the task and in all three conditions at a value of $CV_2 \approx 0.8$ that is matched in monkey M2 (Fig. S3 c). This rules out that the reduction in FF could be partially caused by an increase in firing regularity due to the moderate increase in average firing rate. As an additional control and to rule out the influence of an increasing firing rate on the dynamic FF reduction, we estimated the mean-matched FF as suggested in[45] obtaining the same results (Fig. S2).

Motor cortex model provides mechanistic explanation for task-dependent movement encoding and variability dynamics. We now examine whether and how an attractor-based model of the motor cortex can reproduce our experimental observations in the behaving monkey. To this end we propose a functional spiking neural network model that combines the E/I-clustered topology with an additional decoder module to support behavioral decision making as schematically shown in Fig. 5 d. The core of this model consists of six E/I-clusters, one for each target direction. The network parameters (Table 2) were chosen to match the spontaneous experimental firing rates and baseline FF as estimated before PS.

In our behavioral task simulations we mimicked the behavioral monkey experiment

by applying the same stimulus protocol. Starting with the onset of the PS (Fig. 5 a) we stimulated all excitatory neurons of either one, two or three clusters with the same constant stimulus current throughout the preparatory period to mimic visual cue information. With the RS, i.e. at the end of the preparatory period, stimulation was maintained for only a single cluster that represented the final target. The optimal stimulus amplitude for replicating the experimental results in M1 was determined as 0.1 pA (0.05 pA in M2, Fig. S3) and thus lies in the very low range of tested amplitudes in Fig. 4. In a continuous simulation the model was faced with 150 trials in each of the three stimulus conditions.

We subsequently repeated the exact same analyses of directional encoding and of variability dynamics for the spike train recordings during the model simulation as for the experimental *in vivo* recordings (Fig. 5). The classification analyses, now applied to the entire population of excitatory model neurons, resulted in average decoding accuracies that fully resemble those observed in the experimental data (compare Fig. 5 b and e), recovering the optimal decoding scores of 1, 1/2 and 1/3 that reflect the respective target uncertainty in conditions 1, 2 and 3. This can be explained by the probabilistic nature of a switching activation between the stimulated clusters. As an example, Fig. 6 a shows the spiking activity of all excitatory neurons in the model during a sample trial of Condition 3. With the onset of the PS a stimulus current was applied to clusters 1, 3 and 4 (counting from the bottom). After cue onset, competition arises between these three clusters. All three clusters now share the same higher probability of being active as compared to the non-stimulated clusters. On average the activation times for all three clusters are thus equal and therefore represent the randomly selected final target with probability 1/3 at each point in time. After RS, only the neurons in cluster 3 that represents the selected final target receive stimulating current input and becomes active.

The time-resolved variability analysis of the model data in Fig. 5 c captures the condition-dependent temporal modulation of the average ΔFF observed *in vivo* for monkey M1 (Fig. 5 f) where the reduction in FF is largest for the single target cue and smallest

for the triple target cue. The average spike time irregularity (CV_2) of the model neurons is essentially constant throughout the trial and independent of the target condition, in full qualitative and quantitative agreement with our experimental results (see Supplementary Material and Fig. S3 for the model comparison to monkey M2). The reflection of target uncertainty in the trial-to-trial variability during the preparatory period is mechanistically explained by our model. Stimulation of a single cluster (Condition 1) makes it very likely that this E/I-cluster becomes activated, while all other clusters are likely to stay inactive for most of the 1 s preparatory period. Across repeated trials of the same single target cue, neurons of the same stimulated cluster exhibit elevated firing rates while all other neurons are likely to show a low firing rate. Hence, the across-trial variability of the spike count is low. When there is competition between two or three stimulated clusters, the neurons of these clusters share the overall activation time and hence the FF will be higher.

A simple model of decision making operating on cluster population activities can explain modulation of behavioral reaction times. Inspired by previous models of perceptual decision making[56–59], the decoder module in Fig. 5 d generates a decision variable associated with each cluster. For each target direction d, a leaky integrator governed by an equation of the form

$$\frac{dI_d(t)}{dt} = -\frac{I_d(t)}{\tau_I} + C_d \tag{1}$$

integrates the instantaneous spike count C_d of the corresponding neuron population and forgets with time constant τ_I, representing the integration of target evidence. The decision variable $DV_d(t)$ is formed as

$$DV_d(t) = \frac{I_d(t)}{\sum_{j=1}^{6} I_j(t)} \tag{2}$$

expressing the probability that target direction d represents the correct choice of movement direction at time t. This is similar to multi-class classification. A behavioral decision

was reached when one of the decision variables crossed the common threshold θ within a 400 ms interval following the RS. Threshold crossings after that period were counted as unsuccessful trials. This is similar to the monkey experiments where the monkey had to react within a short time limit. If a decision variable was already above the threshold at the beginning of the response signal (RS), the decision was counted. The threshold was adjusted such that the average accuracy over all sessions is maximized.

The operation of the decoder is illustrated in Fig. 6 a where the respective decision variables $DV_d(t)$ are superimposed on the neural population activity of each target cluster. In this example trial of Condition 3, the clusters 1, 3 and 4 receive stimulation during the preparatory period while target cluster 3 was randomly chosen as the final target indicated by the RS. At the time of RS onset, cluster 1 is still activated and it takes a short time until the change in stimulus takes effect and a switch occurs that finally activates cluster 3. Subsequently $DV_3(t)$ increases and reaches the threshold with a model reaction time of ~ 400 ms.

Fig. 6 b-c show the distributions of reaction times produced by the model and experimental data in M1, respectively, for each condition. It can be seen that, as in the experimental data, the average model reaction times in Condition 1 were much shorter than in Conditions 2 and 3. In contrast to the experiment, anticipated responses were not penalized in the model. If the decision variable of the correct direction was already above threshold at RS, the trial was counted as successful. In Condition 1 this was frequently the case. The shape of the reaction time histogram for Condition 1 (dark blue in Fig. 6 c) suggests that the monkey displayed a similar behavior. Both the data and the model show on average slightly larger response times in Condition 3 compared to Condition 2. The possibility of having prepared for the wrong direction in the model explains the difference in reaction times between the full information and the ambiguous conditions. The same model was successfully adjusted to account for the behavioral data in M2 with a notably different behavioral strategy (see Supplementary Material and Discussion).

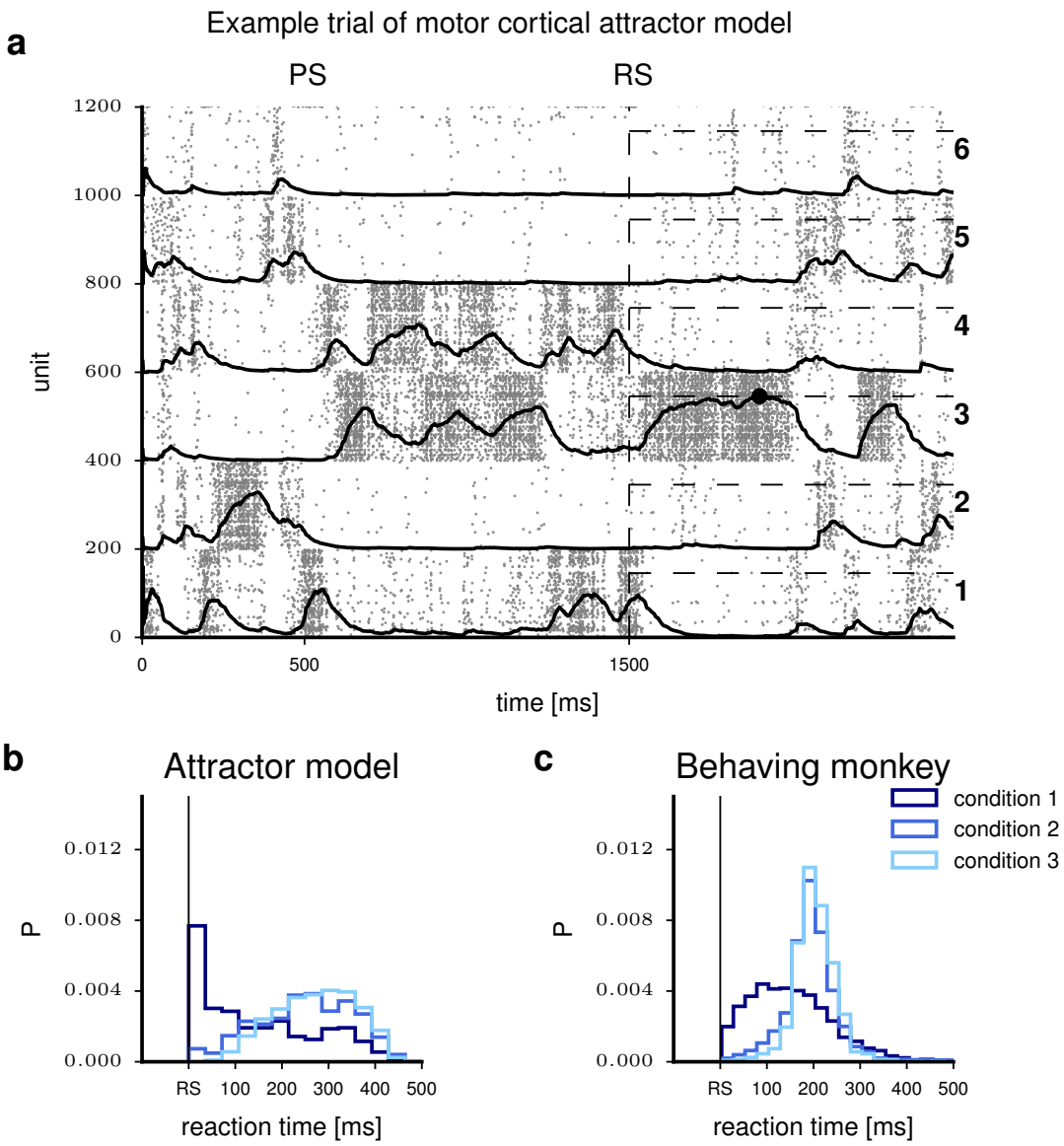

Figure 6: **Decision model generates task-dependent reaction times in monkey M1.** **a**) Raster plot of excitatory cluster activity for an example trial of Condition 3. Solid curves represent the decision variable (DV; equation 2) associated with each cluster. At $t = 1500$ ms the decoder starts integrating, as indicated by the vertical dashed line. Horizontal dashed lines show the level of the decision threshold θ. To the right, target direction numbers are indicated. During the preparatory period (between PS and RS), clusters 1, 3 and 4 are stimulated. The final target (direction 3) was randomly chosen, resulting in a continued stimulation of only cluster 3 after RS. The black circle indicates the decision for a movement in direction 3 by means of threshold crossing. **b-c**) Histograms of reaction times of the model (b) and monkey M1 (c) for all three task conditions.

3 Discussion

Our aim was to construct a functional spiking neural network model for behavioral task performance that can claim biological realism at the level of single neuron spiking statistics within and across trials, and with respect to the generation of realistic behavioral reaction times. We propose a mechanistic model for attractor computation that extends previous spiking attractor network models featuring excitatory clustering[12,16,17,21–23] by adding clustered inhibitory connectivity motivated by accumulating experimental[27,28,30,31,33,34] and theoretical[34,38] evidence for local inhibitory clustering in the neocortex. We show that the proposed architecture ensures local balancing of excitatory and inhibitory synaptic input to individual neurons and supports metastability robustly across a wide range of network parameters, enabling winnerless competition as a signature of cortical attractor computation. The model replicates both, single neuron spiking statistics and behavioral reaction times as measured in two monkeys performing a delayed reaching task in which the degree of prior information about the future movement target is varied. At the neuronal level, we could show that our model can accurately reproduce realistic moderate firing rates, constant spiking irregularity and the task-related dynamics of trial-to-trial spike count variability. In combination with a simple model for evidence integration and decision making, our model can explain the variability of behavioral reaction times in both monkeys. To our knowledge, this is the first model that mechanistically links spiking attractor dynamics to context-dependent modulation of neuronal and behavioral variability, extending on the reaction time model by Mazzucato and colleagues[16]. Taken together, our findings suggest joint excitatory and inhibitory clustering as a biological realistic and powerful mechanism for decision-related computation in the neocortex. We discuss the relevance and limitations of our attractor model and describe future steps of improvement and integration with previously studied models.

The benefits of joint excitatory and inhibitory clustering. Previous studies have considered purely excitatory clustering with a 'blanket of inhibition' to model decision-related activity in the cortex. These have been successful in explaining the trial-to-trial variability observed in *in vivo* recordings[12,16,21–23]. However, in this model architecture metastability can be achieved only in a narrow parameter regime and for small cluster sizes (Fig. 2), which strongly limits its use for functional networks that require robustness against changes in network and input parameters, e.g. due to homeostatic regulation. During cluster activation, single neurons exhibit excessive spike rates[38] (Fig. 4) and an unrealistic regular spiking pattern.

By analyzing the stable fixed points of the mean-field equations for binary networks with purely excitatory clustering, Rost et al.[38] could show that in E-clustered networks switching between attractor states is hampered by the high firing rates attained in active clusters that exhibit a strong self-exciting component. Our proposed remedy of assigning an inhibitory population to each cluster by increasing the corresponding E $\rightarrow$ I, I $\rightarrow$ E and I $\rightarrow$ I synaptic strengths provides a robust solution to both problems. Since inhibition is now equally selective, each excitatory population is held in check by its inhibitory counterpart and hence the fixed points of the active and passive clusters move closer together. Due to the balanced excitatory and inhibitory input drive, individual neurons remain in the fluctuation-driven regime across inactive and active cluster states (Fig. 3). This maintains continuous irregular spiking (Fig. 1) and facilitates switching dynamics (Figs. 2 and 3). The proposed E+I network topology of Litwin-Kumar et al.[21] extends the E-clustered networks by assigning an inhibitory cluster to each excitatory cluster, by increasing the unidirectional E $\rightarrow$ I synaptic strengths. While this increases correlations between these populations it lacks the effect of balancing the input to either of the populations. In line with our analysis of the E-clustered network, the E+I topology of Litwin-Kumar et al. lacks the maintenance of the fluctuation-driven regime during all network states, resulting in the same narrow parameter regime (Fig. 3 and supplementary Fig. S1). We have shown

that our model can depict different system behaviors observable in attractor networks. Depending on the clustering strength J_{E+} the network can show no attractor dynamics, winnerless competition or winner-takes-all dynamics (Fig. 2 c), while being in balance and thus reproducing realistic trial-to-trial as well as spike-interval variability. Besides the clustering strength also the stimulus strength can constrain the winnerless competition to a single cluster and thus lead to an effective winner-takes-all scenario. This covers the functionality of previous work on attractor networks for decision-making and proposes a spiking neuronal implementation that reproduces cortical variability. We conclude that the E/I-clustered topology proposed here provides a robust and biologically realistic network architecture for attractor computation in spiking networks.

Spiking irregularity and trial-by-trial variability are important signatures of cortical activity. Single neuron activity within the neocortex exhibits variability across both short and long time scales. On a short time scale, this variability is evident in the irregular spacing of inter-spike intervals, while on a long time scale it manifests as variability in the number of action potentials across repeated experimental trials. Thus, any biologically realistic model of cortical motor control must offer a mechanistic explanation for both forms of variability observed *in vivo*. Our time-resolved analysis of *in vivo* single unit activity in the motor cortex shows that spiking irregularity is constant throughout spontaneous activity and behavioral task performance with average values of $\mathrm{CV}_2 \approx 0.8$ in both monkeys (Figs. 1 and S3). This is in excellent quantitative agreement with a previous report on the CV_2 estimated in three behaving monkeys during an instructed delay period in the motor cortex[46]. While spiking irregularity is generally slightly more regular in motor areas compared to visual and prefrontal areas, the average CV_2 across all areas remains below unity[60]. A series of *in vitro* experiments in rat neocortical slice preparations using somatic noise current injections under controlled stationary input conditions demonstrated the importance of balancing excitatory with inhibitory synaptic input to achieve realistic

in vivo-like spiking statistics[42,61,62]. Notably, only a high ratio of inhibition to excitation ($R_i = 0.9$) was able to achieve a high coefficient of variation (CV) of approximately 0.7, which comes close to our *in vivo* estimates. Additionally, introducing correlations in synaptic input, a phenomenon commonly observed in spiking neural network simulations, could further enhance spiking irregularity[61,63].

In our E/I-clustered network, excitatory synaptic input is fully balanced with inhibitory synaptic input, while all neurons receive an additional small constant current input. By construction, the balance of synaptic input is maintained during inactivated and activated cluster states (Fig. 3). In our behavioral task simulations, the CV_2 thus remained constant in time, independent of firing rate modulations, and throughout spontaneous activity and task-dependent network stimulation with an average $CV_2 \approx 0.8$, closely matching the observed *in vivo* situation (Fig. 5 and S3). For a stationary spiking process, theoretical consideration predicts that the FF assumes a similar value as the squared coefficient of variation with FF $\approx CV_2^2$ (see Methods). This has been confirmed in *in vitro* experiments that used noise current injection under controlled stationary input conditions[42,61]. The FF typically even assumes values smaller than CV_2^2 due to weak negative serial interval correlations[64], again in full accordance with the theoretical prediction (equation 18). In stark contrast, our *in vivo* analyses show that trial-by-trial variability is high with FF≈ 1.8 during spontaneous activity (i.e. before TS) and thus about three times higher than theoretically expected from the observed spiking irregularity (FF$\approx CV_2^2 \approx 0.65$). Moreover, the FF is modulated during trial time and in a task-related manner while the CV_2 stays constant. This discrepancy between *in vivo* observation and theoretical prediction has led to the interpretation that the observed single unit spiking processes are not stationary across trials violating the theoretical assumption[46,65]. In the spiking attractor model, individual clusters can randomly switch between an inactivated and an activated state. Consequently, each neuron belonging to a cluster generates less or more output spikes. Across repeated experimental trial observation, the number of generated

output spikes is thus highly variable. In our E/I-clustered network, a state of increased (decreased) firing rate is the consequence of increased (decreased) fluctuations of the synaptic input current and membrane potential (fluctuation driven regime) while the balance of excitation and inhibition is retained and, thus, spiking irregularity remains constant.

Functional relevance of our attractor model in the experimental context. The E/I-clustered model not only accounts for realistic neuronal variability dynamics but also reproduces functional and behavioral aspects of movement preparation and decision-making: task-related encoding, and variable reaction times matching the experimental observations. In our experimental design, when the monkey has incomplete information during the preparatory period it can only resolve the ambiguity of multiple targets when the RS signal provides full information. This increased reaction times in M1. We observe the same phenomenon in our model, when adjusting the stimulus amplitude to fit neuronal activity in M1. The mechanistic explanation is that in the multiple-target conditions, the clusters compete and the activity switches between them during the preparation period, which aligns with the previous interpretation of a similar experimental tasks[66,67]. When the RS signal resolves the ambiguity, only one of the clusters retains input stimulation. If this cluster has been active at this point in time, the decision variable reaches the threshold faster (short reaction times) than in the case where a switch is required, leading to longer reaction times (Fig. 6). The same mechanistic explanation underlies short and long reaction times in a recent attractor network model describing behavioral reaction times in anticipatory versus unexpected cues depending on pre-stimulus cluster activation[16]. Kaufmann et al.[67] claim that changes of mind are frequent in decision-making tasks. They show that neuronal activity differs between forced and free changes of mind. They find increased reaction times in trials with suddenly forced decisions, which require re-planning. Our attractor model can provide a mechanistic explanation for their observation as a

forced choice requires switching between attractor states and thus additional time for reaching a decision threshold.

Interpretation of differences in monkey behavior. In discussing the differences in monkey behavior, our study aligns with the notion that visual stimuli influence the motor cortex when they are actively processed and relevant to task performance[68] with typical response latencies of 150-200 ms[40,69,70]. We posit that attention plays a crucial role in directing this processing, affecting the motor representation accordingly. Notably, individual differences in attentional strategies were evident between monkeys M1 and M2, as observed during training and experimental data collection. Monkey M1 tended to capitalize on the early target information in Condition 1, exhibiting faster reactions and anticipatory movements (Fig. 6 c), consistent with the model's behavior. In contrast, Monkey M2 adopted a distinct strategy, disregarding the target cue (PS) and relying solely on the RS for movement initiation across all conditions (Fig. S3 d). This suggests potentially reduced attentiveness to the target cue in M2, possibly opting for a simpler strategy of waiting for the definitive RS[71]. Intriguingly, the model with identical network parameters could replicate this behavioral variation by adjusting the stimulus amplitude during the delay period for M2 compared to M1. In summary, our E/I-clustered model not only accurately replicates the behavioral data from both monkeys but also demonstrates remarkable adaptability by capturing individual differences in behavior through only a single parameter adjustment.

Modelling movement preparation, action selection and movement initiation. Our experimental task involves the perception and processing of visual cues with their uncertainty, the following action-selection, movement preparation, and movement execution. Our work here focuses on the action-selection process. In our experiment the monkeys make unrestricted movements, their hand trajectories were not recorded. Our behavioral read-out is thus restricted to press and release of the buttons. We implemented a simple

drift-diffusion decision model[56,72], which allowed us to quantitatively match average behavioral reaction times and to qualitatively match reaction time distributions in the monkey experiments (Fig. 6). We interpret the temporal integration of the population activity (equation 1) as the integration of evidence about a target-specific action. If the decision variable (equation 2) reaches the threshold, the appropriate action is selected and movement initiation is triggered. Our proposed network topology can not produce muscle-control signals and is thus fundamentally different in its aim from other studies that focus on motor preparation and execution. The general class of recurrent neural networks[73–76] and the more constrained case of a balanced networks[77,78] are capable of producing realistic muscle-control signals, which can be used to control kinetic arm models. These results are based on the viewpoint of the motor cortex as a dynamical system. Within these models an initial state is reached due to a specific stimulation followed by a quasi-deterministic relaxation. The muscle-control signal is generated by linear combination of the generated network activity. In these models it remains unclear how and where the decision for a specific movement is formed and how the preparatory signal is generated. Thus, these approaches are complementary to our model and raise future challenges of combining the here presented spiking approach to action-selection with models of motor preparation and execution.

Model limitations and future directions. We found good agreement between our model and the spiking statistics of motor cortical neurons regarding the slower component of task-related changes in trial-by-trial variability during the preparatory period. However, in monkey M1 we observe an initial fast and transient reduction of the FF immediately after PS where the FF assumes a minimum before returning to a constant level (Fig. 5 c) that is not captured by the attractor model (Fig. 5 f). We hypothesize that this fast transient reduction may be caused by spike frequency adaptation[79] (SFA), which acts as a cell-autonomous mechanism of action-potential induced self-inhibition. SFA is ubiquitous

in spiking neurons in vertebrate and invertebrate nervous systems and has been evidenced in neocortical neurons[80–83]. Previous theoretical results demonstrated that SFA regularizes the spiking process and hence reduces trial-to-trial variability. Under stationary input conditions SFA introduces negative serial correlations of inter-spike intervals[84] as observed in *in vivo* and *in vitro* intracellular recordings from cortical neurons under controlled stationary rate conditions[64,85]. In response to an excitatory stimulus, neurons with SFA in spiking simulations of the random balanced network showed a strong and transient reduction of the FF that relaxes with the time-scale of the SFA-inducing outward conductance[86]. In the next step, we plan to introduce SFA to our E/I-clustered model to investigate whether this can additionally explain the observed fast FF dynamics in response to the visual cue stimulus. The fact that the transient effect of SFA on trial-by-trial variability scales with stimulus amplitude[86] may explain why this was not observed in our data recorded from monkey M2 where a lower stimulus current amplitude was required to match the attractor model to the experimental results (Fig. S3).

Our experimental data is recorded after an extensive training period where both monkeys had learned to accurately perform the task with only very few error trials. In our model we, therefore, assumed that the network connectivity has reached a fixed structure. Future work may investigate how the E/I-clustered network could be learned during training of a task while retaining the local balance of excitation and inhibition[36], for instance through spike-timing-dependent functional plasticity combined with selective stimulation[87–92]. To form and recall these clusters in a stable manner over a long time some form of homeostatic mechanism is crucial. Zenke et al.[89] shows that multiple timescales of homeostatic regulation are necessary to form robust and stable clusters that are functionally relevant. Litwin-Kumar et al.[90] investigate homeostatic synaptic mechanisms that act on I→E synapses in combination with an E→E plasticity rule to form clusters that reflect previously experienced stimuli. The inhibitory plasticity in these studies is globally modulated while excitatory neurons form local clusters and are responsible for functional

representations. Models that implement structural plasticity in spiking networks provide an alternative mechanism that could support cluster formation[93,94]. Motor cortical pyramidal neurons typically display a rather broad tuning to movement direction[40,95]. In our model we implemented one cluster per movement target. Each excitatory neuron belongs to a single cluster that receives target-specific input and, thus, individual neurons cannot exhibit a broad directional tuning. To obtain a more realistic directional tuning behavior in single neurons we propose to implement a modified network topology where clusters that represent neighboring targets share stronger mutual connections to match the target topography in the behavioral space[96]. Alternatively or in addition, target specific network input may be more broadly tuned, stimulating, to a different degree, more than one single cluster.

At any point in time, the population activity of neurons in the motor cortex can be described as a vector in the space that is spanned by the individual neurons, and the evolution of this vector can be studied during the trial. Movement preparation following the perception of the movement instructing cue has been linked to the convergence of the neuronal population activity vector to an optimal subspace[44] such that in each single trial an initial condition will be met that is suitable for movement initiation and execution[97,98]. This theory implies that the trial-by-trial variability reduces from a high spontaneous level where neurons largely act independently to a reduced level when the optimal subspace condition is met[44]. This is in line with our observations. The average population firing rate during the fixed delay preparatory period is similar for movements to different directions and in the three different experimental conditions while the actual target directions can still be decoded accurately[40]. In monkey M1 trial-by-trial variability is reduced to a level that reflects the cued target information during the preparatory period (Fig. 5). In experimental Condition 1 where full target information is cued, the neuronal population activity can reach an initial condition that allows to initiate and execute the appropriate movement, which results in faster anticipatory reaction times. In our attractor model,

this is achieved by the activation of a single cluster (Condition 1) and implies that the total neuron population approaches an optimal subspace characterized by the preferential activation of neurons in the target-specific cluster that leads to fast reaction times. In conditions 2 and 3, two or three clusters compete, implying that the dimensionality of the neuronal population activity reduces during the preparatory period, but the optimal subspace that sets the condition for movement initiation and execution can only be reached with full target information, leading to longer reaction times. Monkey M2 displays a different behavioral strategy. It did not perform anticipatory movements in Condition 1 and showed highly similar reaction time distributions in all three conditions. During the preparatory period, the FF shows the same temporal evolution in all three conditions, implying the same slow reduction of dimensionality. In our interpretation, M2 performed movement preparation only after RS and invariably for all three experimental conditions, and thus the optimal subspace will be occupied only after RS. Future studies of attractor network models should aim at a comparative study of the task-dependent evolution of the neuronal activity space between experimental and simulated spiking data in simultaneous recordings from a large population of neurons.

Methods

Experiment

Behavioral task and recordings in the monkey. The monkey experiments were conducted in Alexa Riehle's lab at the CNRS Marseille, France, in accordance to European and French government regulations. Two monkeys, one male (M1) and one female (M2) performed a delayed center-out reaching task, which involved three different task conditions that differed in the amount of initial target information available to the monkey as illustrated in Fig. 5 a[39,40]. The monkey was seated in front of a panel featuring a hexagonal array of touch-sensitive target LEDs and a central LED indicating the starting position. The monkey initiated a trial by touching the central LED (trial start, TS). During the 1 s delay period starting at $t = 500$ ms the preparatory signal (PS) provided either complete or incomplete information about the final movement target and consisted of either a single target LED (Condition 1), two adjacent target LEDs (Condition 2), or three adjacent target LEDs (Condition 3) that lit up in green. At $t = 1500$ ms the response signal (RS) appeared and one of the green target LEDs turned red. This indicated the final movement target and prompted the monkey to move his hand to that target. In Conditions 2 and 3 the final target was randomly chosen among the PS-cued targets, while the other target LEDs went dark. The times of movement onset (MO) and movement end (ME) were recorded and if the monkey touched the correct target LED, the trial was registered as successful and a drop of juice was given as a reward. A premature onset of the behavior before RS led to the abortion of the trial and no reward was given. Only successful trials were analyzed in the present study.

The task conditions of one, two or three possible targets presented during the 1 s preparatory period were executed in blocks. In each block, 150 trials with randomized target directions were carried out so that each of the directions appeared on average 25 times per condition. Note that to obtain the same number of possible trial types in all

conditions, not all possible combinations of directions for the preparatory stimulus were used in Conditions 2 and 3. Since six combinations are possible for Condition 1, only the pairs 1-2, 3-4 and 5-6 were used in Condition 2 and for Condition 3, only two cases occurred (6-1-2, 3-4-5).

Extracellular recordings were obtained with a multielectrode microdrive (Reitböck system; Thomas Recording) to insert transdurally seven independently movable electrodes within the area of the recording chamber, which was positioned over the motor cortex close to the central sulcus in monkey M1 and slightly more anterior in monkey M2 covering in parts primary and dorsal premotor cortex[40]. Online spike sorting resulted in up to seven simultaneously recorded single-unit spike trains[39]. On each recording day, all three conditions were presented to the monkey so that the responses of individual neurons can be compared across conditions.

Model

Spiking network model. Our spiking network model is composed of leaky integrate-and-fire neurons with exponential synaptic currents where the sub-threshold evolution of the membrane potential V is described by the differential equation

$$\frac{dV}{dt} = \frac{-(V - E_L)}{\tau_m} + \frac{I_{syn} + I_x}{C_m}. \tag{3}$$

In the absence of input, the membrane potential decays exponentially to the resting potential E_L with time constant τ_m. The current I_{syn} represents the synaptic input, I_x is an externally injected current and C_m is the membrane capacitance. If the potential reaches the threshold V_{th} a spike is emitted and V is clamped to a reset voltage V_r for an absolute refractory period τ_r. The synaptic current to a neuron i evolves according to

$$\tau_{syn} \frac{dI^i_{syn}}{dt} = -I^i_{syn} + \sum_j J_{ij} \sum_k \delta\left(t - t^j_k\right) \tag{4}$$

where t_k^j is the time of the arrival of the kth spike from presynaptic neuron j and δ is the Dirac delta function.

To facilitate comparison with previous studies that investigated excitatory cluster topologies we here use similar parameters as provided in Litwin-Kumar et al.[21] and Mazzucato et al.[23] (see Table 1). We briefly explain how we derived the main parameters in the following.

Calibration of the balanced state. We follow the same approach as in Rost et al.[38] for the binary networks by requiring that $\sqrt{K}$ excitatory action potentials arriving within a short time suffice to drive the membrane potential from E_L to V_{th} and hence elicit a spike. For that purpose, we need to compute the deflection in the membrane potential caused by a presynaptic spike.

According to equation 4, a spike arriving at $t = 0$ leads to a postsynaptic current of the form

$$I_{psc}(t) = Je^{-t/\tau_{syn}}\,\Theta\,(t) \tag{5}$$

where J and Θ are the synaptic efficacy and step function, respectively. Inserting this into equation 3 and integrating with $V = 0$ at $t = 0$ the postsynaptic potential is obtained:

$$PSP(t) = J\frac{\tau_m\tau_{syn}}{\tau_m - \tau_{syn}}\left(e^{-t/\tau_m} - e^{-t/\tau_{syn}}\right)\Theta\,(t)\,. \tag{6}$$

The maximal deflection PSP_{max} occurs at $t = \frac{log\frac{\tau_{syn}}{\tau_m}}{(1/\tau_m - 1/\tau_{syn})}$. Note that the PSP amplitude depends on the synaptic as well as the membrane time constants and is therefore different for each synapse type (PSP_{max}^{EE}, PSP_{max}^{EI},...). The scale-free weights are then constructed in the same way as for the binary networks (equations 3 to 8 in Rost et al.[38]) but weighted

by the respective PSP amplitudes:

$$j_{EE} = \frac{V_{th} - E_L}{\sqrt{p_{EE} n_E}} \frac{1}{PSP_{max}^{EE}} \tag{7}$$

$$j_{EI} = -g j_{EE} \frac{p_{EE} n_E}{p_{EI} n_I} \frac{PSP_{max}^{EE}}{PSP_{max}^{EI}} \tag{8}$$

$$j_{IE} = \frac{V_{th} - E_L}{\sqrt{p_{IE} n_E}} \frac{1}{PSP_{max}^{IE}} \tag{9}$$

$$j_{II} = -j_{IE} \frac{p_{IE} n_E}{p_{II} n_I} \frac{PSP_{max}^{IE}}{PSP_{max}^{II}} \tag{10}$$

where g is the relative strength of inhibition. The final weights $J_{\alpha\beta}$ are obtained by dividing by $\sqrt{N}$.

Since we are interested in the temporal dynamics of neuronal variability, we modeled external inputs as constant currents to ensure that all variability arises deterministically inside the network rather than stemming from externally generated Poisson input. In analogy to the 'threshold rate' of Brunel[15], the external current I_x is expressed in terms of the current required to reach the threshold in the absence of synaptic input:

$$I_{th} = \frac{V_{th} - E_L}{\tau_m} C_m. \tag{11}$$

A complex interplay exists between the E and I firing rates and the magnitude of the external currents to the populations. The tuning of the injected currents required to obtain the desired firing rates of 3 and 5 spikes per second for the E and I populations, respectively, was therefore achieved by modeling single units with Poissonian inputs mimicking the network input at the target firing rates. The external inputs could then be increased until the modeled units fired on average at the required rates.

Before introducing structured connectivity we first ensured that the network configuration was operating in the asynchronous-irregular (AI) regime. Irregularity was measured using the squared Coefficient of Variation (CV^2) (as explained in the Data analysis section).

Synchrony of measures such as the instantaneous firing rate or the membrane potential in neural networks can be quantified as[99]:

$$\chi = \sqrt{\frac{\sigma_{pop}^2}{\langle \sigma_i^2 \rangle}}. \tag{12}$$

Here σ_{pop}^2 is the variance of the the population average and $\langle \sigma_i^2 \rangle$ is the average over the individual units' variances. The measure gives unity for totally synchronized activity and for asynchronous activity in networks of size N, one expects $\chi \sim \mathcal{O}\left(\frac{1}{\sqrt{N}}\right)$. Since recording all membrane potentials in simulations is computationally expensive, we computed χ on spike counts measured in bins of 20 ms.

It can be seen in Supplemental Fig. S4 that the networks show the usual characteristics of the balanced state. When excitation dominates, synchronous-regular firing near the saturation rate $1/\tau_r$ is observed. The AI state occurs when g is sufficiently large for inhibition to dominate (Supplemental Fig. S4 a). As in the binary network[38], we choose $g = 1.2$, where $\chi = 0.02 \sim 1/\sqrt{N}$ and $CV^2 = 0.73$ (Supplemental Fig. S4 b). The raster plot shows no discernible structure (Supplemental Fig. S4 c) and the average firing rate is low and constant over time (Supplemental Fig. S4 d). The synaptic currents from excitatory and inhibitory inputs and the external current Ix cancel so that the net input fluctuates around zero (Supplemental Fig. S4 e). Hence, the membrane potentials fluctuate at low values and only occasionally reach the threshold to produce a spike (Supplemental Fig. S4 f). The parameters used for all simulations in this section are summarized in Table 1.

E and E/I-clustered networks. We follow the same connectivity scheme that we introduced for binary networks in our previous work[38]. In summary, for the E-clustered networks, we first divide the excitatory population into Q equally sized clusters with uniform connection probability. Subsequently, we potentiate the synaptic connection

within each cluster by a factor J_{E+}, referred to as cluster strength. A value of $J_{E+} = 1$ represents the random balanced network and the larger J_{E+}, the stronger the weights within the formed clusters. To maintain the overall balance, we decrease the weights among units belonging to different clusters by a factor

$$J_{\alpha-} = \frac{Q - J_{\alpha+}}{Q - 1}, \tag{13}$$

where $\alpha = E$.

For the E/I-clustered networks, we divide not only the excitatory population but also the inhibitory population into Q clusters. We introduce distinct excitatory and inhibitory clustering strengths, J_{E+} and J_{I+}. Then we require that each excitatory cluster selectively influences its corresponding inhibitory cluster and vice versa by increasing the corresponding *EI*, *IE* and *II* weights by a factor J_{I+}. Balance is again maintained by rescaling across-cluster weights according to equation 13 for $\alpha = I$. The overall scaling within and across clusters (indicated by the superscripts in/out) is expressed as follows: $J^{in}_{EE} = J_{E+}J_{EE}$, $J^{out}_{EE} = J_{E-}J_{EE}$, $J^{in}_{\alpha\beta} = J_{I+}J_{\alpha\beta}$, $J^{out}_{\alpha\beta} = J_{I-}J_{\alpha\beta}$ for $\alpha\beta \in (EI, IE, II)$. We have shown in Rost et al.[38] that the inhibitory clustering needs to be weaker than the excitatory clustering to obtain realistic firing rates. To capture this relationship, we introduce a proportionality factor R_J, such that

$$J_{I+} = 1 + R_J(J_{E+} - 1), \tag{14}$$

where $0 \leq R_J \leq 1$. A value of $R_J = 1$ implies the same cluster strength for inhibitory and excitatory clusters ($J_{E+} = J_{I+}$) and $R_J = 0$ makes the inhibitory population unclustered ($J_{I+} = 1$, which represents the E-clustered networks). Throughout the current study, we use $R_J = 3/4$ based on our previous results[38] where we showed that this value of R_J can prevent firing rate saturation in up states.

Attractor model of motor cortex for simulation of the behavioral monkey task. We designed a model with six E/I-clusters with 200/50 excitatory/inhibitory neurons each. We adjusted the clustering parameter J_{E+} for the smaller network size to achieve robust metastability under spontaneous network conditions and an average FF that approximates the average experimental value. The external input currents were slightly adapted to obtain spontaneous firing rates of the excitatory neurons matching the average experimental value. The inhibitory neurons were tuned to have slightly higher firing rates. To match the task evoked firing rates and FF the only free parameter is the stimulus amplitude. No additional parameter tuning was performed and all model parameters are listed in Table 2.

In an ongoing simulation we define successive trials of 2 s length. We randomly define the start of the first trial (TS). This is followed after 500 ms by the onset of the PS and after another 1000 ms by the onset of the RS, which lasted for 400 ms. PS and RS were realized as constant stimulating currents I_{stim}. After each trial we draw a random inter-trial interval in the range of $1.5 - 1.7$ s before we start the next trial, to allow the network to relax to its spontaneous state. The variance in this relaxation period was intended to avoid any effects of periodicity. For each task condition the model executed 150 trials and for analysis the trials were cut from this long continuous simulation. All analyses of model spiking data were performed identically to the analyses of the *in vivo* spiking data. Our setup implies that each neuron in our model is sharply selective to only a single target direction.

The threshold θ applied to the decision variable DV (equation 2) was adjusted to maximize the performance of the model. The time constant of integration τ_I was set to 50 ms which represents an intermediate value between very fast reactions directly when the threshold is reached at RS and very slow integration where the threshold was not reached during the RS interval.

Data analysis

Quantifying neural variability. The FF quantifies the dispersion of spike counts for single neurons across repeated observations of the same experimental condition (trials)

$$FF = \frac{\sigma_c^2}{\mu_c}, \tag{15}$$

where σ_c^2 and μ_c are the variance and mean count over trials. The estimation of the FF is biased towards unity for small estimation windows. However, this bias quickly becomes negligible when the estimation window is several multiples of the mean inter-spike interval (ISI)[42,65].

Interval statistics are usually characterized by the Coefficient of Variation (CV) of the empirical ISI distribution,

$$\mathrm{CV} = \frac{\sigma_{ISI}}{\mu_{ISI}}. \tag{16}$$

Here, σ_{ISI} and μ_{ISI} are the standard deviation and mean of the intervals between action potentials. Estimating the CV requires some caution, as modulations in firing rate increase the interval variability. Another problem with estimating the CV follows from finite-size estimation windows. In an estimation window of width T, only $ISIs < T$ can be observed. If the underlying process has non-zero probabilities for larger intervals, the CV will be underestimated; this effect is known as right-censoring[42,100]. For estimating the interval variability we therefore used the CV_2[41,43], which is computed for all pairs of consecutive intervals as:

$$\mathrm{CV}_2 = 2\left\langle \frac{|\tau - \tau'|}{\tau + \tau'} \right\rangle. \tag{17}$$

Here, $\langle ... \rangle$ denotes averaging and τ and τ' are consecutive ISIs. The CV_2 is assumed to be largely unaffected by firing rate changes on time scales that are slower than two consecutive intervals.

A neuron generates action potentials and a spike train may be considered a realization

of a stochastic point process[101]. We may thus interpret interval and count variability in the light of stochastic point process theory. For any defined point process the expectation values for the FF and CV can be computed and are related. For stationary processes, i.e. when the point process definition is fix and the point process intensity (i.e. the firing rate in neuronal terms) is constant, we obtain[64,65,102,103]

$$\lim_{T\to\infty} \mathsf{FF} = \mathrm{CV}_\infty^2 \Big(1 + 2\,\xi\Big) \qquad \text{with} \quad \xi = \sum_{i=1}^{\infty} \xi_i \tag{18}$$

where ξ denotes the ith-order linear correlation coefficient for pairs of intervals (ISI_k, ISI_{k+i}). In practical terms, the limit notation reflects the estimation biases (see above) for finite observation windows of length T. In the case of a renewal process where the stochastic point process is defined by the inter-spike interval distribution and, thus, no serial correlations of inter-spike intervals exist, we expect FF$\approx$ CV^2 for long T, for the special case of the Poisson process this simplifies to FF$=$ CV^2 $=1$ irrespective of the length T of the observation interval.

In the present study we consider an experimental trial design where in each trial the monkeys (and our models) perform a given behavioral task. For our analyses we grouped trials that belong to the same experimental condition and with identical preparatory target cue (PS) and final target cue (RS) presentation. For each single unit we then count the number of spikes in each trial and within the defined observation window to compute the FF across a group of trials. For the same observation window we computed the per-trial average CV_2 before averaging across trials. Using a sliding observation window allows to observe the task-related temporal dynamics of the average FF and CV_2. To allow a fair comparison of variability statistics across conditions, additional precautions were taken. We required that units had at least ten spikes in the 2 s interval after trial start and that at least ten trials were recorded per condition and cued target direction. To enable the comparison across experimental conditions, we only included units and cued

target directions where those criteria were met in all conditions. Under the assumption of stationarity we expect the relation given in equation 18 for FF and CV_2. Strong deviations from this relation indicate non-stationarity across trials[42,65].

Decoding of movement direction. To assess how much directional information is contained in the population activity we reproduced the approach of Rickert et al.[40] and constructed pseudo-populations of all available units as follows: the data set was divided into five groups and each group in turn served as the test set while the model was trained on the remaining groups, and the average score of the five models was calculated. At each point in time and using five-fold cross-validation we computed the decoding accuracy of the logistic regression classifier as the fraction of correctly predicted movement directions averaged over all six different directions as

$$\text{Decoding accuracy} = \frac{1}{C}\sum_{c=1}^{C}\frac{N^c_{correct}}{N^c_{total}}. \tag{19}$$

Parameter	Unit	Value
N	-	$4000(E), 1000(I)$
E_L	mV	0
V_{th}	mV	15
V_R	mV	0
C_m	pF	1
τ_m	ms	$20(E), 10(I)$
τ_{syn}	ms	$3(E), 2(I)$
τ_r	ms	5
p_{EE}	-	0.2
p_{EI}, p_{IE}, p_{II}	-	0.5
g	-	1.2
J_{EE}	pA	0.25
J_{EI}	pA	-0.66
J_{IE}	pA	0.19
J_{II}	pA	-1.00
I_x	pA	$2.13I_{th}(E), 1.24I_{th}(I)$

Table 1: Standard network parameters for the simulations of the E/I-clustered spiking neural network simulations.

Parameter	Unit	Value
N	-	$1200(E), 300(I)$
E_L	mV	0
V_{th}	mV	15
V_R	mV	0
C_m	pF	1
τ_m	ms	$20(E), 10(I)$
τ_{syn}	ms	$3(E), 2(I)$
τ_r	ms	5
p_{EE}	-	0.2
p_{EI}, p_{IE}, p_{II}	-	0.5
g	-	1.2
J_{EE}	pA	0.45
J_{EI}	pA	-1.20
J_{IE}	pA	0.34
J_{II}	pA	-1.83
I_x	pA	$1.25I_{th}(E), 0.78I_{th}(I)$
Q	-	6
J_{E+}	-	3.2
R_J	-	3/4

Table 2: Network parameters for the simulation of the monkey task.

Acknowledgments

This project was funded by the German Research Foundation (DFG), in parts through the Collaborative Research Center 'Motor Control in Health and Disease' (DFG-SFB 1451, Project A06, ID 431549029 to M.P.N.) and under the Institutional Strategy of the

University of Cologne within the German Excellence Initiative (DFG-ZUK 81/1 to M.P.N. and S.A). S.A. received funding through European Union's Horizon 2020 Framework Programme for Research and Innovation under grant agreement number 945539 (Human Brain Project SGA3). F.S. was funded through the DFG Research Training Group Neural Circuit Analysis (DFG-RTG 1960, grant no. 233886668 to M.P.N.). We thank Dr. Moritz Deger for fruitful discussions and support in the early phase of the project. We thank Jörn Rickert for the monkey drawing.

Author contribution statement

Conceptualization, V.R., T.R., A.R., and M.P.N.; experimental task design and experiments, A.R.; data analysis, V.R., T.R., F.S. and M.P.N.; computational modeling, V.R., T.R. and F.S.; writing – original draft, V.R. and M.P.N.; writing – review & editing, V.R., F.S., S.A., A.R., M.P.N., funding acquisition, S.A. and M.P.N.

Data availability

All experimental data and obtained simulation results have been deposited at G-Node (`https://gin.g-node.org/nawrotlab/EI_clustered_network`).

Code availability

All code to reproduce the figures, including simulation code and code to analyze the experimental and simulated data are available at `https://github.com/nawrotlab/ClusteredNetwork_pub`.

References

1. Renart, A. & Machens, C. K. Variability in neural activity and behavior. *Current opinion in neurobiology* **25**, 211–20 (2014).

2. Sakai, K. & Miyashita, Y. Neural organization for the long-term memory of paired associates. *Nature* **354**, 152–155 (1991).

3. Wills, T. J., Lever, C., Cacucci, F., Burgess, N. & O'Keefe, J. Attractor dynamics in the hippocampal representation of the local environment. *Science* **308**, 873–876 (2005).

4. Knierim, J. J. & Zhang, K. Attractor Dynamics of Spatially Correlated Neural Activity in the Limbic System. *Annual Review of Neuroscience* **35**, 267–285 (2012).

5. Kopec, C. D., Erlich, J. C., Brunton, B. W., Deisseroth, K. & Brody, C. D. Cortical and subcortical contributions to short-term memory for orienting movements. *Neuron* **88**, 367–377 (2015).

6. Miconi, T., McKinstry, J. L. & Edelman, G. M. Spontaneous emergence of fast attractor dynamics in a model of developing primary visual cortex. *Nature Communications* **7**, 1–10 (2016).

7. Pereira, U. & Brunel, N. Attractor Dynamics in Networks with Learning Rules Inferred from In Vivo Data. *Neuron* **99**, 227–238.e4 (2018).

8. Inagaki, H. K., Fontolan, L., Romani, S. & Svoboda, K. Discrete attractor dynamics underlies persistent activity in the frontal cortex. *Nature* **566**, 212–217 (2019).

9. Finkelstein, A. *et al.* Attractor dynamics gate cortical information flow during decision-making. *Nature neuroscience* **24**, 843–850 (2021).

10. Stroud, J. P. *et al.* Effects of noise and metabolic cost on cortical task representations. *bioRxiv* (2023). https://www.biorxiv.org/content/early/2023/12/08/2023.07.11.548492.full.pdf.

11. Hopfield, J. J. Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences of the United States of America* **79**, 2554–2558 (1982).

12. Amit, D. & Brunel, N. Model of global spontaneous activity and local structured activity during delay periods in the cerebral cortex. *Cerebral Cortex* **7**, 237–252 (1997).

13. Roudi, Y. & Latham, P. E. A Balanced Memory Network. *PLoS Computational Biology* **3**, e141 (2007).

14. van Vreeswijk, C. & Sompolinsky, H. Chaotic balanced state in a model of cortical circuits. *Neural computation* **10**, 1321–71 (1998).

15. Brunel, N. Dynamics of Sparsely Connected Networks of Excitatory and Inhibitory Spiking Neurons. *Journal of Computational Neuroscience* **8**, 183–208 (2000).

16. Mazzucato, L., La Camera, G. & Fontanini, A. Expectation-induced modulation of metastable activity underlies faster coding of sensory stimuli. *Nature Neuroscience* **22**, 787–796 (2019).

17. La Camera, G., Fontanini, A. & Mazzucato, L. Cortical computations via metastable activity. *Current Opinion in Neurobiology* **58**, 37–45 (2019). 1906.07777.

18. Daie, K., Svoboda, K. & Druckmann, S. Targeted photostimulation uncovers circuit motifs supporting short-term memory. *Nature neuroscience* **24**, 259–265 (2021).

19. Mazzucato, L. Neural mechanisms underlying the temporal organization of naturalistic animal behavior. *eLife* **11**, e76577 (2022).

20. Recanatesi, S., Pereira-Obilinovic, U., Murakami, M., Mainen, Z. & Mazzucato, L. Metastable attractors explain the variable timing of stable behavioral action sequences. *Neuron* **110**, 139–153.e9 (2022).

21. Litwin-Kumar, A. & Doiron, B. Slow dynamics and high variability in balanced cortical networks with clustered connections. *Nature Neuroscience* **15**, 1498–1505 (2012).

22. Deco, G. & Hugues, E. Neural Network Mechanisms Underlying Stimulus Driven Variability Reduction. *PLoS Computational Biology* **8**, e1002395 (2012).

23. Mazzucato, L., Fontanini, A. & La Camera, G. Dynamics of multistable states during ongoing and evoked cortical activity. *The Journal of neuroscience* **35**, 8214–31 (2015).

24. Packer, A. M. & Yuste, R. Dense, unspecific connectivity of neocortical parvalbumin-positive interneurons: a canonical microcircuit for inhibition? *The Journal of neuroscience : the official journal of the Society for Neuroscience* **31**, 13260–71 (2011).

25. Fino, E. & Yuste, R. Dense inhibitory connectivity in neocortex. *Neuron* **69**, 1188–1203 (2011).

26. Albantakis, L. & Deco, G. Changes of mind in an attractor network of decision-making. *PLOS Computational Biology* **7**, 1–13 (2011).

27. Xue, M., Atallah, B. V. & Scanziani, M. Equalizing excitation–inhibition ratios across visual cortical neurons. *Nature* **511**, 596–600 (2014).

28. Lee, S.-H. *et al.* Parvalbumin-Positive Basket Cells Differentiate among Hippocampal Pyramidal Cells. *Neuron* **82**, 1129–1144 (2014).

29. Morishima, M., Kobayashi, K., Kato, S., Kobayashi, K. & Kawaguchi, Y. Segregated excitatoryinhibitory recurrent subnetworks in layer 5 of the rat frontal cortex. *Cerebral Cortex* **27**, 5846–5857 (2017).

30. Arkhipov, A. *et al.* Visual physiology of the layer 4 cortical circuit in silico. *PLoS Computational Biology* **14**, e1006535 (2018).

31. Khan, A. G. *et al.* Distinct learning-induced changes in stimulus selectivity and interactions of GABAergic interneuron classes in visual cortex. *Nature Neuroscience* **21**, 851–859 (2018).

32. Znamenskiy, P. *et al.* Functional selectivity and specific connectivity of inhibitory neurons in primary visual cortex. *bioRxiv* 294835 (2018).

33. Shin, M. *et al.* Both excitatory and inhibitory neurons transiently form clusters at the outermost region of the developing mammalian cerebral neocortex. *Journal of Comparative Neurology* **527**, 1577–1597 (2019).

34. Najafi, F. *et al.* Excitatory and Inhibitory Subnetworks Are Equally Selective during Decision-Making and Emerge Simultaneously during Learning. *Neuron* **105**, 165 – 179.e8 (2020).

35. Znamenskiy, P. *et al.* Functional specificity of recurrent inhibition in visual cortex. *Neuron* (2024).

36. Hennequin, G., Agnes, E. J. & Vogels, T. P. Inhibitory Plasticity: Balance, Control, and Codependence. *Annual Review of Neuroscience* **40**, 557–579 (2017).

37. Okun, M. & Lampl, I. Instantaneous correlation of excitation and inhibition during ongoing and sensory-evoked activities. *Nature Neuroscience* **11**, 535–537 (2008).

38. Rost, T., Deger, M. & Nawrot, M. P. Winnerless competition in clustered balanced networks: inhibitory assemblies do the trick. *Biological Cybernetics* 1–18 (2017).

39. Bastian, A., Schoner, G. & Riehle, A. Preshaping and continuous evolution of motor cortical representations during movement preparation. *European Journal of Neuroscience* **18**, 2047–2058 (2003).

40. Rickert, J., Riehle, A., Aertsen, A., Rotter, S. & Nawrot, M. P. Dynamic Encoding of Movement Direction in Motor Cortical Neurons. *Journal of Neuroscience* **29**, 13870–13882 (2009).

41. Holt, G. R., Softky, W. R., Koch, C. & Douglas, J. Comparison of discharge variability in vitro and in vivo in cat visual cortex neurons. *Journal of Neurophysiology* **75**, 1806–1814 (1996).

42. Nawrot, M. P. *et al.* Measurement of variability dynamics in cortical spike trains. *Journal of neuroscience methods* **169**, 374–90 (2008).

43. Ponce-Alvarez, A., Kilavik, B. E. & Riehle, A. Comparison of local measures of spike time irregularity and relating variability to firing rate in motor cortical neurons. *Journal of Computational Neuroscience* **29**, 351–365 (2010).

44. Churchland, M. M., Yu, B. M., Ryu, S. I., Santhanam, G. & Shenoy, K. V. Neural variability in premotor cortex provides a signature of motor preparation. *The Journal of neuroscience : the official journal of the Society for Neuroscience* **26**, 3697–712 (2006).

45. Churchland, M. M. *et al.* Stimulus onset quenches neural variability: a widespread cortical phenomenon. *Nature Neuroscience* **13**, 369–378 (2010).

46. Riehle, A., Brochier, T., Nawrot, M. P. & Grün, S. Behavioral Context Determines Network State and Variability Dynamics in Monkey Motor Cortex. *Frontiers in Neural Circuits* **12**, 52 (2018).

47. Churchland, A. K. *et al.* Variance as a signature of neural computations during decision making. *Neuron* **69**, 818–831 (2011).

48. Isaacson, J. & Scanziani, M. How Inhibition Shapes Cortical Activity. *Neuron* **72**, 231–243 (2011).

49. Froemke, R. C. Plasticity of Cortical Excitatory-Inhibitory Balance. *Annual Review of Neuroscience* **38**, 195–219 (2015).

50. Van Vreeswijk, C. & Sompolinsky, H. Chaos in neuronal networks with balanced excitatory and inhibitory activity. *Science* **274**, 1724–1726 (1996).

51. Anderson, J. S., Carandini, M. & Ferster, D. Orientation Tuning of Input Conductance, Excitation, and Inhibition in Cat Primary Visual Cortex. *Journal of Neurophysiology* **84**, 909–926 (2000).

52. Zhang, L. I., Tan, A. Y. Y., Schreiner, C. E. & Merzenich, M. M. Topography and synaptic shaping of direction selectivity in primary auditory cortex. *Nature* **424**, 201–205 (2003).

53. Dehghani, N. *et al.* Dynamic Balance of Excitation and Inhibition in Human and Monkey Neocortex. *Scientific Reports* **6**, 23176 (2016).

54. Rubin, R., Abbott, L. F. & Sompolinsky, H. Balanced excitation and inhibition are required for high-capacity, noise-robust neuronal selectivity. *Proceedings of the National Academy of Sciences of the United States of America* **114**, E9366–E9375 (2017).

55. Nawrot, M. P., Aertsen, A. & Rotter, S. Single-trial estimation of neuronal firing rates: From single-neuron spike trains to population activity. *Journal of Neuroscience Methods* **94**, 81–92 (1999).

56. Gold, J. I. & Shadlen, M. N. The neural basis of decision making. *Annual Review of Neuroscience* **30**, 535–574 (2007).

57. Ratcliff, R. & McKoon, G. The diffusion decision model: theory and data for two-choice decision tasks. *Neural computation* **20**, 873–922 (2008).

58. Drugowitsch, J. & Pouget, A. Probabilistic vs. non-probabilistic approaches to the neurobiology of perceptual decision-making. *Current opinion in neurobiology* **22**, 963–969 (2012).

59. Meckenhäuser, G., Krämer, S., Farkhooi, F., Ronacher, B. & Nawrot, M. P. Neural representation of calling songs and their behavioral relevance in the grasshopper auditory system. *Frontiers in Systems Neuroscience* **8**, 183 (2014).

60. Mochizuki, Y. *et al.* Similarity in neuronal firing regimes across mammalian species. *Journal of Neuroscience* **36**, 5736–5747 (2016).

61. Stevens, C. F. & Zador, A. M. Input synchrony and the irregular firing of cortical neurons. *Nature neuroscience* **1**, 210–217 (1998).

62. Harsch, A. & Robinson, H. P. Postsynaptic variability of firing in rat cortical neurons: the roles of input synchronization and synaptic nmda receptor conductance. *Journal of Neuroscience* **20**, 6181–6192 (2000).

63. Nawrot, A. *et al.* Embedding living neurons into simulated neural networks. In *First International IEEE EMBS Conference on Neural Engineering, 2003. Conference Proceedings.*, 229–232 (IEEE, 2003).

64. Nawrot, M. P. *et al.* Serial interval statistics of spontaneous activity in cortical neurons in vivo and in vitro. *Neurocomputing* **70**, 1717–1722 (2007).

65. Nawrot, M. P. Analysis and interpretation of interval and count variability in neural spikes trains. In Grün, S. & Rotter, S. (eds.) *Analysis of Parallel Spike Trains*, 34–58 (Springer Verlag, New York, Berlin, 2010).

66. Cisek, P. & Kalaska, J. F. Neural correlates of reaching decisions in dorsal premotor cortex: Specification of multiple direction choices and final selection of action. *Neuron* **45**, 801–814 (2005).

67. Kaufman, M. T., Churchland, M. M., Ryu, S. I. & Shenoy, K. V. Vacillation, indecision and hesitation in moment-by-moment decoding of monkey motor cortex. *eLife* **4**, e04677 (2015).

68. Kwan, H., Mackay, W., Murphy, J. & Wong, Y. Properties of visual cue responses in primate precentral cortex. *Brain Research* **343**, 24–35 (1985).

69. Riehle, A., Kornblum, S. & Requin, J. Neuronal correlates of sensorimotor association in stimulus–response compatibility. *Journal of Experimental Psychology: Human Perception and Performance* **23**, 1708 (1997).

70. Kaufman, M. T. *et al.* The largest response component in the motor cortex reflects movement timing but not movement type. *eneuro* **3** (2016).

71. Riehle, A. & Requin, J. Monkey primary motor and premotor cortex: single-cell activity related to prior information about direction and extent of an intended movement. *Journal of Neurophysiology* **61**, 534–549 (1989). PMID: 2709098, `https://doi.org/10.1152/jn.1989.61.3.534`.

72. Ding, L. & Gold, J. I. The basal ganglia's contributions to perceptual decision making. *Neuron* **79**, 640–649 (2013).

73. Sussillo, D., Churchland, M. M., Kaufman, M. T. & Shenoy, K. V. A neural network

that finds a naturalistic solution for the production of muscle activity. *Nature Neuroscience* **18**, 1025–1033 (2015).

74. Elsayed, G. F., Lara, A. H., Kaufman, M. T., Churchland, M. M. & Cunningham, J. P. Reorganization between preparatory and movement population responses in motor cortex. *Nature Communications* **7**, 13239 (2016).

75. Michaels, J. A., Dann, B. & Scherberger, H. Neural population dynamics during reaching are better explained by a dynamical system than representational tuning. *PLOS Computational Biology* **12**, 1–22 (2016).

76. Russo, A. A. *et al.* Motor cortex embeds muscle-like commands in an untangled population response. *Neuron* **97**, 953–966.e8 (2018).

77. Hennequin, G., Vogels, T. P. & Gerstner, W. Optimal control of transient dynamics in balanced networks supports generation of complex movements. *Neuron* **82**, 1394–1406 (2014).

78. Kao, T.-C., Sadabadi, M. S. & Hennequin, G. Optimal anticipatory control as a theory of motor preparation: A thalamo-cortical circuit model. *Neuron* **109**, 1567–1581.e12 (2021).

79. Benda, J. & Herz, A. V. A universal model for spike-frequency adaptation. *Neural computation* **15**, 2523–2564 (2003).

80. Fuhrmann, G., Markram, H. & Tsodyks, M. Spike frequency adaptation and neocortical rhythms. *Journal of neurophysiology* **88**, 761–770 (2002).

81. Lundstrom, B. N., Higgs, M. H., Spain, W. J. & Fairhall, A. L. Fractional differentiation by neocortical pyramidal neurons. *Nature neuroscience* **11**, 1335–1342 (2008).

82. Ha, G. E. & Cheong, E. Spike frequency adaptation in neurons of the central nervous system. *Experimental neurobiology* **26**, 179 (2017).

83. Salaj, D. *et al.* Spike frequency adaptation supports network computations on temporally dispersed information. *Elife* **10**, e65459 (2021).

84. Farkhooi, F., Muller, E. & Nawrot, M. P. Adaptation reduces variability of the neuronal population code. *Physical Review E* **83**, 050905 (2011).

85. Engel, T. A., Schimansky-Geier, L., Herz, A. V., Schreiber, S. & Erchova, I. Subthreshold membrane-potential resonances shape spike-train patterns in the entorhinal cortex. *Journal of neurophysiology* **100**, 1576–1589 (2008).

86. Farkhooi, F., Froese, A., Muller, E., Menzel, R. & Nawrot, M. P. Cellular adaptation facilitates sparse and reliable coding in sensory pathways. *PLoS computational biology* **9**, e1003251 (2013).

87. Vogels, T. P., Sprekeler, H., Zenke, F., Clopath, C. & Gerstner, W. Inhibitory plasticity balances excitation and inhibition in sensory pathways and memory networks. *Science* **334**, 1569–1573 (2011).

88. Ocker, G. K., Litwin-Kumar, A. & Doiron, B. Self-Organization of Microcircuits in Networks of Spiking Neurons with Plastic Synapses. *PLoS Computational Biology* **11** (2015).

89. Zenke, F., Agnes, E. J. & Gerstner, W. Diverse synaptic plasticity mechanisms orchestrated to form and retrieve memories in spiking neural networks. *Nature Communications* **6**, 6922 (2015).

90. Litwin-Kumar, A. & Doiron, B. Formation and maintenance of neuronal assemblies through synaptic plasticity. *Nature Communications* **5**, 5319 (2014).

91. Mongillo, G., Rumpel, S. & Loewenstein, Y. Inhibitory connectivity defines the realm of excitatory plasticity. *Nature Neuroscience* **21**, 1463–1470 (2018).

92. Wu, Y., Hengen, K. B., Turrigiano, G. G. & Gjorgjieva, J. Homeostatic mechanisms regulate distinct aspects of cortical circuit dynamics. *bioRxiv* 790410 (2019).

93. Bourjaily, M. A. & Miller, P. Excitatory, inhibitory, and structural plasticity produce correlated connectivity in random networks trained to solve paired-stimulus tasks. *Frontiers in Computational Neuroscience* **5**, 37 (2011).

94. Gallinaro, J. V., Gašparović, N. & Rotter, S. Homeostatic control of synaptic rewiring in recurrent networks induces the formation of stable memory engrams. *PLOS Computational Biology* **18**, 1–40 (2022).

95. Georgopoulos, A. P., Kalaska, J. F., Caminiti, R. & Massey, J. T. On the relations between the direction of two-dimensional arm movements and cell discharge in primate motor cortex. *Journal of Neuroscience* **2**, 1527–1537 (1982).

96. Hennequin, G., Ahmadian, Y., Rubin, D. B., Lengyel, M. & Miller, K. D. The dynamical regime of sensory cortex: Stable dynamics around a single stimulus-tuned attractor account for patterns of noise variability. *Neuron* **98**, 846–860.e5 (2018).

97. Afshar, A. *et al.* Single-Trial Neural Correlates of Arm Movement Preparation. *Neuron* **71**, 555–564 (2011).

98. Meirhaeghe, N., Riehle, A. & Brochier, T. Parallel movement planning is achieved via an optimal preparatory state in motor cortex. *Cell Reports* **42**, 112136 (2023).

99. Golomb, D. & Hansel, D. The number of synaptic inputs and the synchrony of large, sparse neuronal networks. *Neural Computation* **12**, 1095–1139 (2000).

100. Wiener, M. C. An adjustment to the time-rescaling method for application to short-trial spike train data. *Neural computation* **15**, 2565–76 (2003).

101. Tuckwell, H. *Introduction to Theoretical Neurobiology: Volume 2, Nonlinear and Stochastic Theories*. Cambridge Studies in Advanced Mathematics (Cambridge University Press, 1988).

102. Cox, D. & Lewis, P. *The Statistical Analysis of Series of Events*. Methuen's Monographs on Applied Probability and Statistics (Springer Netherlands, 1966).

103. Chacron, M. J., Longtin, A. & Maler, L. Negative interspike interval correlations increase the neuronal capacity for encoding time-dependent stimuli. *Journal of Neuroscience* **21**, 5328–5343 (2001).