Excitable dynamics of NREM sleep: a unifying model for neocortex and hippocampus

Daniel Levenstein[12], *György Buzsáki*[12], *John Rinzel*[13]*

SUMMARY

During non-rapid eye movement (NREM) sleep, the neocortex and hippocampus continuously alternate between states of neuronal spiking and inactivity. To study the mechanisms of NREM alternation dynamics, we used a mean field model of a recurrent adapting neural population. Analysis of the model reveals how the interplay between recurrence, adaptation, and drive can result in four distinct regimes of alternation dynamics. By directly comparing model dynamics with experimental observations of NREM sleep, we find that the neocortical dynamics match the model in a regime in which a stable active state is interrupted by transient inactive states (slow waves) while hippocampal dynamics match the model in a regime in which a stable inactive state is interrupted by transient active states (sharp waves). We propose that during NREM sleep, hippocampal and neocortical populations are excitable: each in a stable state from which noise or perturbation can evoke the stereotyped population events that support NREM functions.

Keywords: NREM sleep, UP/DOWN states, Sharp Wave-Ripples

[1] Center for Neural Science, New York University
[2] New York University Neuroscience Institute
[3] Courant Institute for Mathematical Sciences, New York University
* Lead Contact, Correspondence: rinzeljm@gmail.com

INTRODUCTION

NREM sleep is dominated by a neocortical "slow oscillation" (Steriade et al., 1993) in which neuronal populations alternate between periods of spiking activity (UP states) and periods of hyperpolarization (DOWN states). Neocortical DOWN states correspond to large deflections of the local field potential (LFP) termed "slow waves" (or "delta waves", (Ball et al., 1977; Buzsáki et al., 1988)), which are the signature of mammalian NREM sleep (Achermann and Borbely, 1997). The slow oscillation supports the consolidation of recently learned memories by temporally coupling to hippocampal sharp wave-ripple events (SWRs) (Siapas and M. A. Wilson, 1998; Sirota et al., 2003). SWRs are themselves the hippocampal analog of neocortical UP states: periods of spiking (SWR) separated by periods of relative inactivity (inter-SWR) (Buzsáki, 2015). In spite of the different durations of active/inactive states in the two regions, the hippocampus and neocortex are both cortical tissue and are under similar neuromodulatory influence during NREM sleep (Lee and Dan, 2012). Thus, it is possible that the neocortical slow oscillation and hippocampal SWRs can be explained by similar principles.

Experiments in multiple physiological contexts have revealed that SWRs and UP/DOWN alternations are a locally-generated "default state" of hippocampal and neocortical tissue (Buzsáki, 2006; Sanchez-Vives et al., 2017). Beyond NREM, they are observed during quiet wakefulness, (Buzsáki et al., 1983; Petersen et al., 2003) under anesthesia (Steriade et al., 1993; Ylinen et al., 1995), and during *in vitro* slice or culture preparations (Colgin et al., 2004; Sanchez-Vives and McCormick, 2000). Understanding how alternation dynamics are generated, and how they differ between these physiological contexts, is important to understand their role in cortical function. Early modeling work demonstrated that recurrent excitatory connections are sufficient to support bistability, a condition in which a neural population can exist in a DOWN or an UP state (Latham et al., 2000; H. R. Wilson and Cowan, 1972), between which noise can induce stochastic alternations between the two states (Jercog et al., 2017). Subsequent experiments found that neuronal adaptation - activity-driven negative feedback in cortical excitatory cells - is involved in UP/DOWN alternations (Contreras et al., 1996; Sanchez-Vives and McCormick, 2000). Biophysical spiking models (Bazhenov et al., 2002; Compte et al., 2003; Destexhe, 2009; Hill and Tononi, 2005) and population rate models (Parga and Abbott, 2007) were then used to demonstrate that alternations between UP and DOWN states can also arise from a variety of adaptation mechanisms. However, it is not yet clear how the relative contributions of adaptation and recurrence influence UP/DOWN dynamics in a population.

Models with recurrence and adaptation have been directly matched to neocortical UP/DOWN dynamics during anesthesia and in slice preparations (Curto et al., 2009; Jercog et al., 2017; Mattia and Sanchez-Vives, 2012). However, recurrent adapting models have not yet been matched to data from NREM sleep in naturally sleeping animals. Furthermore, these studies have had differing results: the UP/DOWN alternations in slice were found to be adaptation-mediated oscillations (Mattia and Sanchez-Vives, 2012), while those in the anesthetized animal were found to reflect noise-induced switches between bistable states (Jercog et al., 2017). Thus, the mechanisms of UP/DOWN alternations during NREM remain unclear.

To study alternation dynamics during NREM sleep, we used an idealized model of an adapting recurrent neuronal population. The model is able to produce four distinct regimes of alternation dynamics, which encompass those seen in previous models and in different physiological contexts. Analysis of the model reveals how the relative contribution of recurrent excitation and adaptation, as well as the level of drive, determine the regime of UP/DOWN alternations in a neuronal population. By directly matching the durations of modeled and experimental UP/DOWN states, we find that neocortical and hippocampal dynamics during NREM sleep are neither oscillations nor bistable, but are *excitable*: each population rests in a stable state from which suprathreshold fluctuations can induce a transient population event that is terminated by the influence of adaptation. Specifically, the neocortex maintains a stable UP state with occasional transitions to a transient DOWN state (slow waves) while the hippocampus is in a stable DOWN state with occasional transitions to a transient UP state (SWRs). Under the influence of noise, these excitable regimes produce UP/DOWN alternations with the asymmetric duration distributions seen during NREM sleep. We further observe that the stability of the rodent neocortical UP state changes over the course of sleep in a way that resembles the well-characterized stages of NREM sleep in humans (Berry et al., 2017). Our findings reveal a unifying picture of hippocampal and neocortical dynamics during NREM sleep, with implications for NREM function.

RESULTS

In agreement with previous findings, we observed large amplitude slow waves in the neocortical LFP during NREM sleep in naturally-sleeping rats (Figure 1A). Population spike rate, calculated in overlapping 40ms windows, showed alternations between low spike rate UP states (~1-2Hz/cell) and DOWN states of spiking inactivity (<0.1Hz/cell), coincident with the LFP slow waves (Figure 1B). We used the coincidence of large-amplitude peaks in the low-frequency LFP (0.5 Hz - 8 Hz) with drops in the power of the high-frequency band of the LFP (100 Hz - 400 Hz; representative of spiking activity, (Watson et al., 2017)) to identify DOWN and UP states (Figure S1, Methods). The durations of UP and DOWN states both showed skewed distributions (Figure 1C). However, the statistics of UP and DOWN states were highly asymmetrical: UP state durations were longer (mean = 1.7±0.92s) and more variable (CV = 1.1±0.27) than those of DOWN states (mean = 0.21±0.05s, CV = 0.38±0.06).

We also observed alternation between active and inactive states in a separate dataset of hippocampal population activity. Short duration (50-100ms) strongly synchronous spiking was coincident with hippocampal sharp-wave ripples (SWRs) and alternated with inter-SWR intervals with sparse spiking (Figure 1E,F). We used a previously described method to detect SWR events as the coincidence of increase ripple-band power and a low-frequency sharp wave (see Methods, (Grosmark and Buzsáki, 2016)). Similar to the neocortical UP/DOWN states, the durations of SWRs and interSWR periods also showed skewed distributions, and were highly asymmetrical. InterSWR durations were much longer (mean = 2.0±0.22s) and more variable (CV = 1.3±0.10) than those of SWR events (mean = 0.06±0.005s, CV = 0.33±0.04 (Figure 1G).

To simulate the alternation dynamics, we used a mean field model for the spike rate of an adapting neuronal population. Using the observed duration distributions, we were able to match our dynamic model to the UP/DOWN dynamics observed during NREM sleep in the neocortex (Figure 1D) and SWR dynamics in the hippocampus (Figure 1H). However, we also found that the model embodies a rich repertoire of physiologically-relevant dynamical regimes that extend beyond those seen during NREM sleep. In the following sections, we first introduce the model and its dynamic regimes (Figure 2). We then describe an analysis of the model that exposes how the interplay between recurrent excitation, adaptation, and drive can account for a range of UP/DOWN dynamics in neural populations (Figure 3-4). Next, we quantitatively match the model dynamics to those seen during NREM sleep (Figure 5-6). Finally, we add an inhibitory population to the model and show how balanced excitation and inhibition can account for the low firing rates during neocortical UP states and features of UP/DOWN transitions (Figure 7).

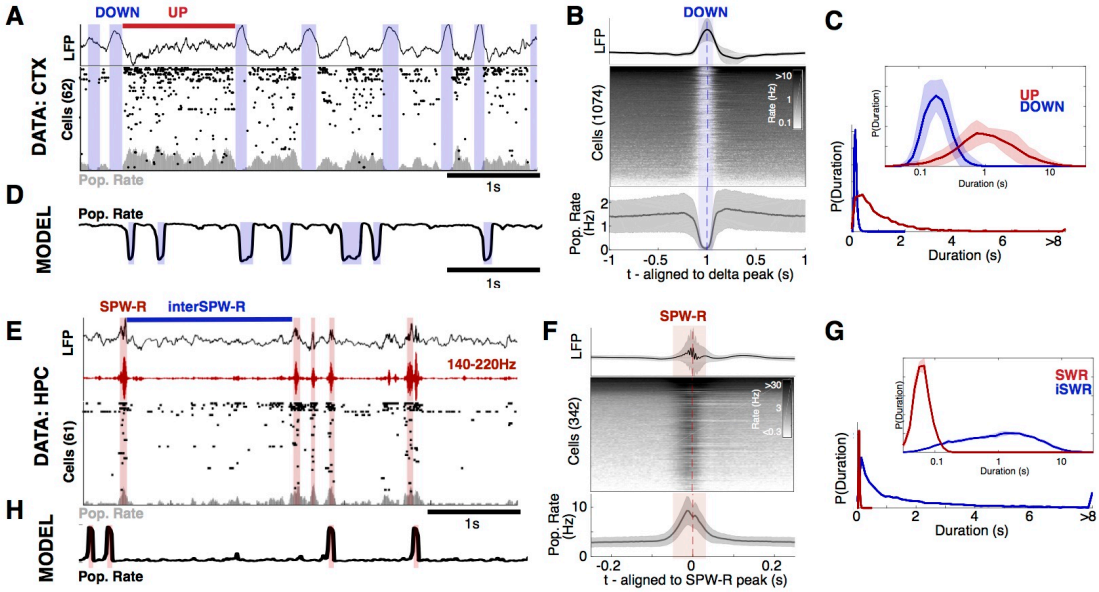

Figure 1 Neocortical UP/DOWN and hippocampal SPW-R dynamics during NREM sleep. **A:** A sample of data from rat mPFC during NREM sleep. Data was collected using high-density silicon probes (Watson et al 2016). LFP and spike times from cortical neurons were extracted as reported previously (see Methods). Neocortical slow waves are coincident with population-wide non-spiking DOWN states (B), which alternate with UP states of longer duration (C) These dynamics are mimicked by a population rate model (D). **B:** Peri-event time histogram aligned to delta peaks for the LFP (top), all recorded cells (middle), and population rate (bottom). Mean +/- standard deviation over all recordings shown. **C:** UP/DOWN state dwell time distribution (bottom) in linear (example recording) and logarithmic scale (mean +/-std over all recordings). **D:** Simulation of the model (eqns 1-2). Parameters determined by matching in vivo and simulated UP/DOWN state dwell times, as described in section: *Neocortex is in an Excitable$_{UP}$ regime during NREM sleep*. **E:** NREM LFP from a representative hippocampal recording (Grosmark and Buzsaki 2016). Detected sharp wave-ripples (SPW-R) state indicated with red. **F:** Peri-event time histogram aligned to SPW-R peaks for the LFP (top), all recorded cells (middle), and population rate (bottom). Mean +/- standard deviation over all recordings shown. **G:** SPW-R and inter-SPW-R duration distributions from all recordings. **H:** Simulated *r-a* model with the best-matching parameters, as described in section: *Hippocampus is in an Excitable$_{DOWN}$ regime during NREM sleep*.

UP/DOWN dynamics in an adapting excitatory population model

In the spirit of Wilson and Cowan (H. R. Wilson and Cowan, 1972), we model cortical activity in terms of the mean firing rate of a neuronal population, here an excitatory population with rate $r(t)$, subject to a slow negative feedback (i.e. adaptation), $a(t)$ (Figure 2A, see Methods for details and parameters).

$$\tau_r \frac{dr}{dt} = -r + R_\infty\big(wr - ba + I + \xi(t)\big) \qquad (1)$$

$$\tau_a \frac{da}{dt} = -a + A_\infty(r) \qquad (2)$$

In the model, rate and adaptation each vary between an inactive state ($r = 0, a = 0$) and an active state ($r = 1, a = 1$). Equations 1-2 describe how r and a evolve in time as a function of the net input to the population: the sum of the recurrent excitation with weight w and a background level of drive with a tonic parameter I, and noisy fluctuations $\xi(t)$, minus adaptation weighted by gain parameter b. Under constant net input, the population rate will approach a steady state level given by the input-output relation, $R_\infty(input)$. Adaptation is similarly activated by neuronal activity to a steady state $A_\infty(r)$. $R_\infty(input)$ and $A_\infty(r)$ are taken to be sigmoidal functions. The time constants τ_r and τ_a determine how quickly $r(t)$ and $a(t)$ will approach their steady state values and time has been non-dimensionalized to arbitrary model units (AU) so $\tau_r = 1$.

Model dynamics can be represented as a trajectory in the r-a phase plane (Figure 2B, Supplemental video 1, (Strogatz, 2014)). In the phase plane, trajectories are predictable from the r- and a-nullclines: two curves defined by the conditions $\frac{dr}{dt} = 0$ and $\frac{da}{dt} = 0$. Steady states, or fixed points, of activity are found at intersections of the nullclines and may be either stable (attractors) or unstable. The r-nullcline is N-shaped with a right branch at $r{\sim}1$ and a left branch at $r{\sim}0$, which correspond to UP and DOWN states of activity (Figure 2B). When adaptation is slow (i.e. $\tau_a \gg \tau_r$), trajectories move horizontally toward the UP/DOWN branches with time scale τ_r. At the UP/DOWN branches, trajectories drift along the r-nullcline as adaptation activates or inactivates with the time scale τ_a. If the branch contains a stable fixed point, the system will remain in the UP or DOWN state until some perturbation induces a transition to the opposing branch. If there is no fixed point, the trajectory transitions to the opposing branch at the turning point of the r-nullcline (Figure 2B). In this way, an UP or DOWN state in the model can be either stable: requiring a perturbation to evoke a transition to the opposite state, or

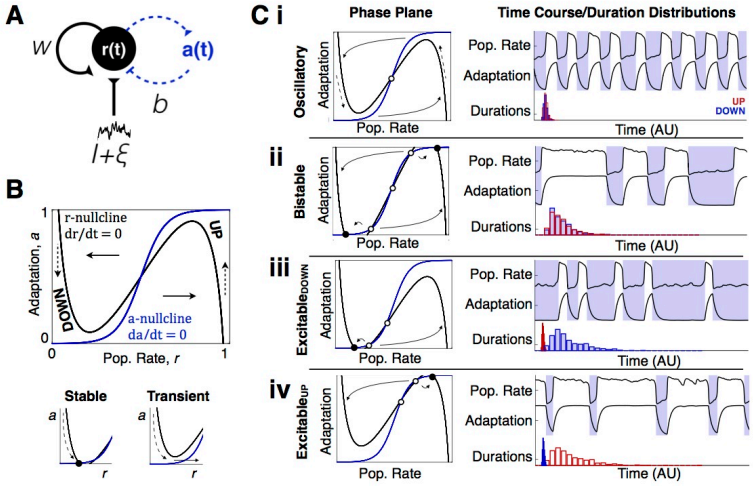

Figure 2: UP/DOWN dynamics in an adapting recurrent neural population. **A:** Wilson-Cowan model for a neural population with slow adaptive process. **B:** r-a Phase plane. Model dynamics are seen as trajectories in the phase plane that follow equations 1-2. Dashed arrows indicate slow vertical trajectories at timescale of τ_a, solid arrows indicate fast horizontal trajectories at timescale of τ_r. Nullclines (dr/dt = 0, da/dt = 0) and their intersections graphically represent dynamics for a set of parameter values. Left and right branches of the r-nullcline correspond to DOWN and UP states, respectively. Stable UP/DOWN states are seen as stable fixed points at nullcline intersections. Transient UP/DOWN states are seen as r-nullcline branch with no intersection. **C:** Four UP/DOWN regimes available to the model, as distinguished by location and stability of fixed points. Representative phase plane (Left), simulated time course and UP/DOWN state duration distributions (right, time units arbitrary) for each regime. Stable fixed points are represented by filled circles, unstable fixed points by empty circles. Parameters: (i-iv) b=1, (i,iii,iv) w=6 (ii) w=6.3, (i) I=2.5 (ii)

transient: automatically transitioning to the opposite state due to the activation or inactivation of adaptation.

Depending on parameter values, the nullclines can take one of four configurations that correspond to distinct regimes of UP/DOWN dynamics (Figure 2C). In the presence of noise, one regime corresponds to noisy oscillations (i) and three correspond to noise-induced alternation patterns (ii-iv). Comparable regimes are found in similar idealized models (Jercog et al., 2017; Mattia and Sanchez-Vives, 2012), which reinforces the generality of the model dynamics.

In the oscillatory regime (Figure 2Ci), activity alternates between transient UP and DOWN states at a relatively stable frequency. Adaptation activates during the UP state and brings the population to the DOWN state, during which adaptation inactivates and the population returns to the UP state. The oscillation corresponds to a limit cycle trajectory in the phase plane, surrounding a single, unstable fixed point. Because $r(t)$ is fast compared to the slow adaptation, the $r(t)$ time course and the phase plane trajectory are square-shaped, with rapid transitions between UP and DOWN states (Supplemental video 1, Figure S2).

In the case of two stable fixed points, the system is in a bistable regime (Figure 2Cii). In this regime, adaptation is not strong enough to induce UP/DOWN state transitions. However, sufficiently large (suprathreshold) fluctuations can perturb the population activity to cross the middle branch of the r-nullcline, resulting in a transition to the opposing branch. Thus, the presence of noise induces alternations between stable UP and DOWN states, resulting in highly variable UP/DOWN state durations.

In the case of a single stable fixed point, the system can still show UP/DOWN alternations in one of two excitable regimes. If the stable fixed point is on the DOWN branch (Figure 2Ciii), the system is in an $\text{Excitable}_{\text{DOWN}}$ regime. The population will remain in the stable DOWN state in the absence of any external influence. However, a brief activating input to the population can trigger a rapid transition to a transient UP state, during which adaptation activates, leading to a return to the DOWN branch. In the presence of noise, UP states are triggered spontaneously by net activating fluctuations. The time course of the model in the $\text{Excitable}_{\text{DOWN}}$ regime shows long DOWN states of variable durations punctuated by brief stereotyped UP states.

Conversely, if the stable fixed point is on the UP branch, the system is in an $\text{Excitable}_{\text{UP}}$ regime (Figure 2Civ, Supplemental Video 2). Brief inactivating input can elicit a switch from the stable UP state to a transient DOWN state, during which adaptation deactivates, leading to a

return to the UP branch. Thus, in the presence of noise, DOWN states are triggered spontaneously by net-inactivating fluctuations. The time course will show longer UP states of variable durations with stereotypically brief DOWN states. These two regimes (Figure 2Ciii,iv) are excitable because relatively small fluctuations in population rate can "excite" the population out of a stable steady state and induce disproportionately large, stereotyped, population events: a transient UP state in the case of the Excitable$_{DOWN}$ regime and a transient DOWN state in the case of the Excitable$_{UP}$ regime, followed by a return to steady state.

Recurrence, adaptation, and drive control UP/DOWN regimes

How do properties of a population determine dynamical regime? We next use numerical and analytical methods from dynamical systems theory (Strogatz, 2014) to create maps of the UP/DOWN regimes in parameter space (e.g., Figure 3C, 3F, see Methods) and reveal how intrinsic and network properties determine the properties of UP/DOWN dynamics in our model. We find that, given sufficient recurrent excitation, the dynamical regime of UP/DOWN alternations is determined by two factors: the level of drive, and the relative strength of recurrence and adaptation. We then analyze model simulations to reveal how UP/DOWN state durations act as a signature of dynamical regime, allowing us to later match modeled and experimental dynamics.

UP/DOWN alternations are possible only if the population can potentially exist in an UP or a DOWN state. This requires adequate strength of recurrent excitation, w, to self-maintain the UP state under conditions of low drive. We show this first for a reduced case without adaptation dynamics ($b = 0$). In this case, the population rate $r(t)$ satisfies

$$\frac{dr}{dt} = -r + R_{\infty}(wr + I)$$

The phase space is reduced from a plane to a line, and dynamics of the population rate correspond to motion along the r-axis following eqn. 3 (Figure 3A). The motion is rightward (rate increasing) where $dr/dt > 0$ and leftward where $dr/dt < 0$. If recurrent excitation is sufficiently strong, the graph of dr/dt vs. r is N-shaped and there can be two stable (and one unstable) fixed points: an UP state of high activity at $r \approx 1$ and a DOWN state of low activity at $r \approx 0$.

The population rate at fixed points, r_{ss}, depends on the level of drive, I, as is described by the *effective* input/output relation (I/O curve) of the recurrently connected population (Figure 3B). If recurrence is weak ($w = 0$), the I/O curve increases monotonically with I. With increased

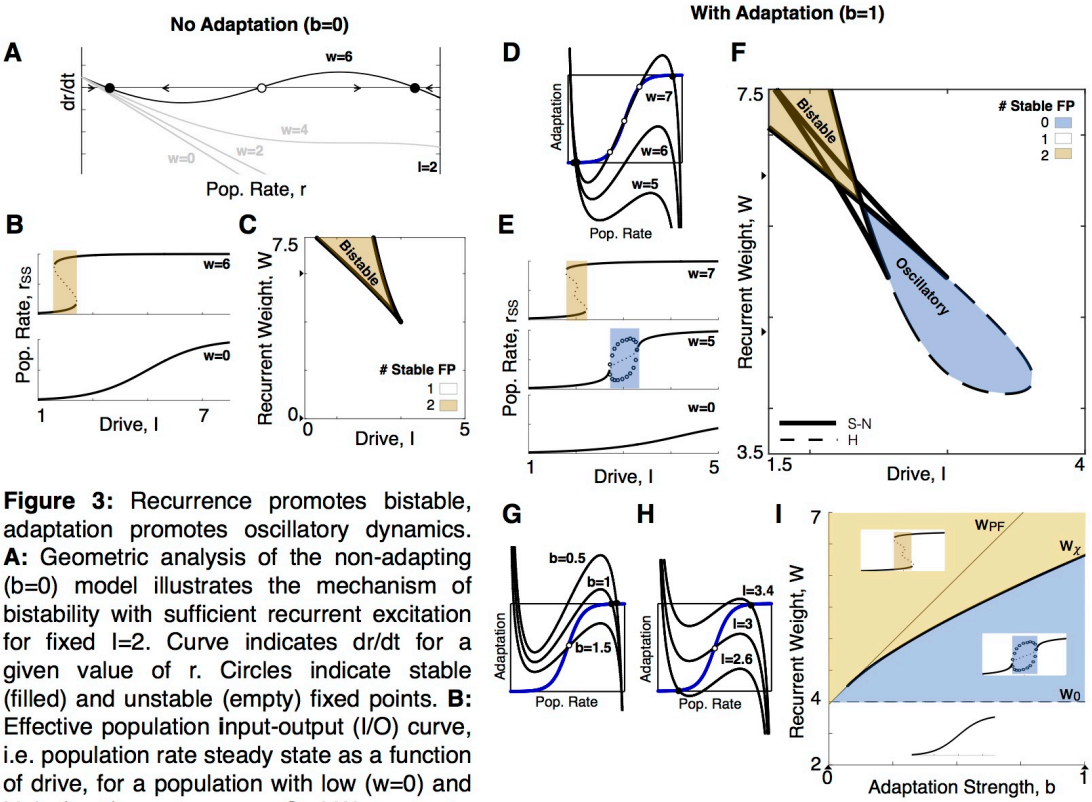

Figure 3: Recurrence promotes bistable, adaptation promotes oscillatory dynamics. **A:** Geometric analysis of the non-adapting (b=0) model illustrates the mechanism of bistability with sufficient recurrent excitation for fixed I=2. Curve indicates dr/dt for a given value of r. Circles indicate stable (filled) and unstable (empty) fixed points. **B:** Effective population input-output (I/O) curve, i.e. population rate steady state as a function of drive, for a population with low (w=0) and high (w=6) recurrence. **C:** I-W parameter space, bistable region appears in a cusp bifurcation at high recurrence and low input. **D:** Effect of recurrence on the r-nullcline (b=1, I=2). **E:** Effective I/O curves for populations with increasing levels of recurrent excitation. Solid lines indicate stable fixed points, dotted lines indicate unstable fixed points, and circles indicate the upper/lower bounds of a limit cycle. Yellow indicates bistable regime (2 stable fixed points), blue indicates oscillatory regime (0 stable fixed points with a stable limit cycle). **F:** I-W Bifurcation diagram reveals the "Butterfly Catastrophe" motif, yellow indicates bistability, blue indicates oscillations. Solid line, saddle-node (S-N) bifurcations; dashed line, Hopf (H) bifurcation. **G:** Effect of adaptation strength, b, on r-nullcline (w=6, I=2.75). **H:** Effect of drive, I, on r-nullcline (w=5, b=1). **I:** Regions in the w-b parameter plane for distinct I/O properties, indicated by insets. Blue shading indicates parameter values for a oscillatory-centered I/O curve, yellow shading indicates a bistable-centered I/O curve. Boundary curves correspond to parameters where bifurcations vs I coalesce, see Methods.

recurrence, the I/O curve shows a central region of bistability between a low-rate fixed point at weak drive and a high-rate fixed point at strong drive. In the I-w parameter space (Figure 3C), the bistable region (yellow) has borders that correspond to saddle-node bifurcations at the knees of the I/O curve. UP/DOWN bistability emerges at a critical value of recurrence ($w = 4$), at a level of drive for which the unconnected population would be minimally activated ($I = I_0 - 2$, where $R_\infty(I_0) = 0.5$, see Methods). Consequently, a first general insight of recurrent population rate models is that UP/DOWN bistability will emerge in neuronal populations with sufficiently strong recurrent excitation, during conditions of low drive.

With adaptation dynamics reintroduced ($b = 1$), the mechanism for UP/DOWN bistability is like that in the adaptation-free case. Increasing recurrence (larger w) enhances the N-shape of the r-nullcline, allowing for multiple stable fixed points at intersections of the r- and a-nullclines (Figure 3D). The effective I/O curve is again bistable-centered for strong recurrence (Figure 3E, top). The region of I-w parameter space with multiple fixed points is now "butterfly-shaped" (Figure 3F, S2), with a bistable regime again inside the yellow region at higher values of w.

In contrast to the effect of w, stronger adaption (larger b) diminishes the N-shape of the r-nullcline and decreases likelihood for multiple fixed points (Figure 3G). As a result, the network can now oscillate at intermediate values of w, for which the I/O curve appears oscillatory-centered (Figure 3E, middle). The oscillatory region in I-w parameter space is blue, with borders corresponding to Hopf bifurcations (Figure 3F). Increasing the strength of adaptation increases the domain of this oscillatory region in I-w parameter space (Figure S2).

Increasing drive raises the r-nullcline and brings the population from a stable DOWN state at low drive to a stable UP state at high drive, with fixed points that trace out the I/O curve (Figure 3H). Due to their opposite effects on the r-nullcline, adaptation and recurrence oppositely influence the dynamic regime at the I/O curve's center region (Figure 3I). For low levels of recurrence, the I/O curve increases monotonically with a stable fixed point for each I-value and no UP/DOWN alternations are possible. At a critical value of recurrence, UP/DOWN alternations emerge at $I_{1/2}$: the axis of symmetry of the I/O curve (Figure S3, Methods). When adaptation is very weak, only bistability is possible and UP/DOWN alternations emerge in a cusp bifurcation as in Figure 3C. With sufficient adaption UP/DOWN alternations emerge in Hopf bifurcation as in Figure 3F (See methods for analytical solutions). With sufficient recurrent

excitation, the population will have a bistable-centered I/O curve if recurrence is stronger (Figure 3I, yellow) or an oscillatory-centered I/O curve if adaptation is stronger (Figure 3I, blue). Thus, the relative strength of recurrence and adaptation defines a spectrum from bistable-centered to oscillatory-centered response properties.

In the absence of noise or external perturbation, only the oscillatory regime will alternate between UP and DOWN states. Noise extends the parameter domain for UP/DOWN alternations beyond the oscillatory regime by enabling transitions out of stable fixed points. Consider the effects of noise on the oscillatory-centered I/O curve (Figure 4A). Within the oscillatory regime the simulated population rate alternates regularly between transient UP and DOWN states, and UP/DOWN state durations reflect the time scale of adaptation, $\sim\tau_a$ (Figure S4). For I-values above the oscillatory regime, noise can evoke transitions from the stable UP state to a transient DOWN state (an Excitable$_{UP}$ regime). DOWN state durations still reflect the time scale of adaptation, τ_a, but UP state durations are much greater than τ_a. They now reflect the waiting time for random fluctuations to drop the system out of the UP state attractor, and thus UP state durations vary with noise amplitude (Figure S4). For *I*-values further above the oscillatory regime, the effective stability of the UP state increases; DOWN states are less frequent because the larger fluctuations needed to end the UP state are less frequent. Thus, UP states become progressively longer as I is increased, while DOWN states stay approximately the same duration ($\sim\tau_a$). The same case is seen for values of I below the oscillatory regime but with UP/DOWN roles reversed (i.e. an Excitable$_{DOWN}$ regime).

The effective I/O relations, accounting for noise, are represented by statistical properties of UP/DOWN state durations. The duration distributions plotted vs. drive form a crossed-pair, with a center symmetrical portion (i.e. an oscillatory (Figure 4A) or bistable (Figure S4) regime) flanked by the asymmetrical Excitable$_{DOWN}$ and Excitable$_{UP}$ regimes. A complementary view of the I/O properties is the graph of decreasing fraction of time in the DOWN state as drive is increased (Figure 4Aii) This Prob(DOWN), or silence density, has been used as an experimental metric of the degree of cortical synchronization (Mochol et al., 2015). By this terminology, more synchronized regimes correspond to more time spent in the DOWN state, and less synchronized regimes correspond to more time spent in the UP state. Increased drive eventually leads to an UP-only (asynchronous) regime. Thus, the level of drive defines a spectrum from more synchronized to less synchronized dynamics in the model which, along with the spectrum

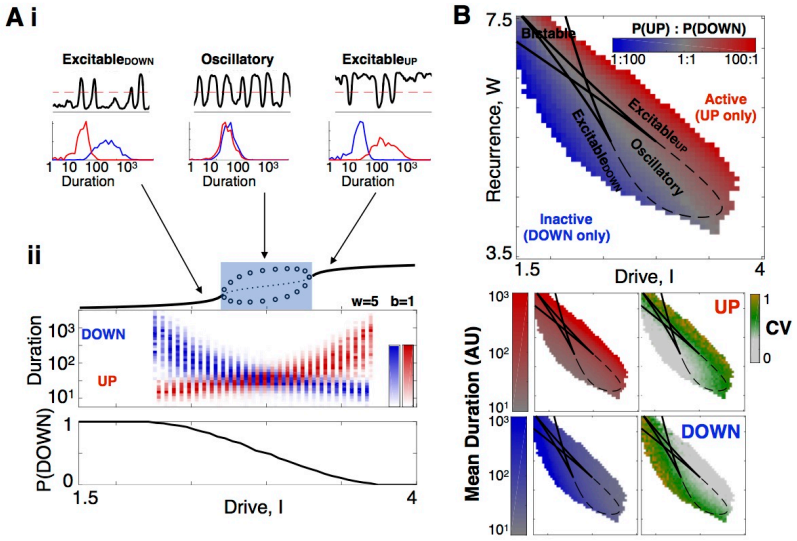

Figure 4: Drive defines a spectrum from DOWN-dominated to UP-dominated dynamics **A:** Population I/O curve reveals spectrum from more synchronized (lower drive) to less synchronized (higher drive) dynamics. **i:** Simulated time course (r vs t, top) and dwell time distributions (bottom) for low (I=2.6), intermediate (I=3), and high (I=3.4) levels of drive. ii**:** Numeric fixed point solutions (top, from Fig 3E), UP/DOWN state duration distributions (middle), and silence density (bottom) as a function of drive to the system. **B:** Duration statistics as a signature of UP/DOWN regime. Mean and CV of simulated dwell times for the I-W parameter space, as in 3F.

from bistable to oscillatory dynamics described above, defines a set of characteristic axes in the "space" of UP/DOWN alternation dynamics.

In sum, the UP/DOWN state duration statistics reflect the underlying dynamical regime. The mean durations vary continuously over the parameter plane as UP/DOWN durations increase/decrease with the level of drive (Figure 4B and correspondingly in the I-b space, Figure S4C). However, the duration variability (as measured by the coefficient of variation, CV) shows sharp transitions at the boundaries between regimes, which reflect the different nature of transitions out of stable and transient states. In general, the durations of stable states are longer and more variable while those of transient states are shorter and less variable, effectively distinguishing oscillatory, bistable, and excitable dynamics. In the following sections, we use the duration distributions to match experimentally-observed alternation dynamics with specific dynamic regimes in the model.

Neocortex is in an Excitable$_{UP}$ regime during NREM sleep

The durations of neocortical UP/DOWN states as presented in Figure 1 are indicative of an Excitable$_{UP}$ regime in our model. Neocortical UP states during NREM are longer (1.7±0.92s) compared to DOWN states (0.21±0.05s), and more irregular (CV_{UP} = 1.1±0.27; CV_{DOWN} = = 0.38±0.06) (Figure 5A). We directly compared the simulated and experimentally-observed dynamics by matching the statistics of experimental UP/DOWN durations to those in Figure 4D. We found that the region of model parameter space in which the CV_{UP}, CV_{DOWN} and ratio of means is within 2 standard deviations of the experimental durations is in the Excitable$_{UP}$ regime (Figure 5B, red outline). We next compared the shapes of the duration distributions between model and experiment. For each model realization (i.e. each point in the I-w parameter plane), we calculated the similarity between simulated and experimental duration distributions for each recording session in the experimental dataset (Figure S5, Methods). We found high similarity for each session over a substantial domain in the I-w plane, indicating that the UP/DOWN dynamics of the model provide a good match for those observed during NREM in vivo (Figure S5). The domain of high similarity between animal data and the model fell in the Excitable$_{UP}$ regime, as indicated by the 25 best fit points and in the average values of similarity (over all 25 sessions) in I-w parameter space (Figure 5B) and in the I-b parameter space (Figure S6D). The simulated time course (Figure 5D) and duration distributions (Figure 5C) using the parameter set with highest mean similarity over all sessions revealed a good match between experimental and modeled dynamics. We thus found that NREM sleep in the rodent neocortex is

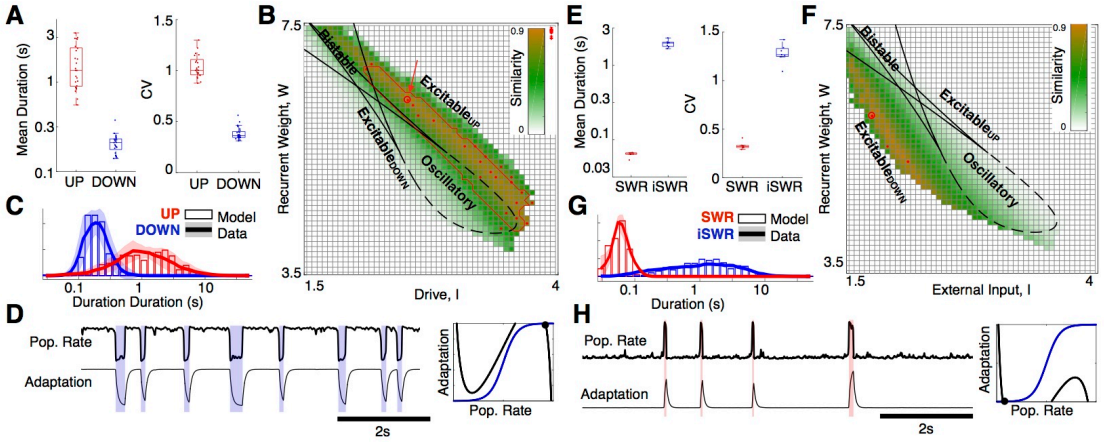

Figure 5 Duration matching reveals that neocortex (A-D) and hippocampus (E-H) are in excitable regimes during NREM sleep. **A:** Mean and CV of UP/DOWN state durations, each point is a single recording session. **B:** Similarity between simulated and experimental duration distributions (see Figure S5). Color indicates mean similarity over all recordings, red line outlines region for which model simulations fall within mean+/-2STD of experimentally observed CV_{UP}, CV_{DOWN} and $mean_{UP}/mean_{DOWN}$ ratio. 25 red dots indicate most similar parameters for each of the 25 recordings, and large circle (arrow) is best mean similarity, see Methods. **C:** Neocortical duration distributions for data (mean+/-std over all recordings) and model simulation in the best mean similarity parameters. **D:** Simulated r-a model with the best-matching parameters indicated in panel B. (I=2.64, W=6.28, b=1). (Right) Phase plane diagram for model with best fit parameters, in the $Excitable_{UP}$ regime. **E:** Mean and CV of SPW-R/interSPW-R state durations, each point is a single recording session. **F:** Similarity between simulated and experimental duration distributions as in panel B. **G:** Hippocampal Duration distributions for data (mean +/-std over all recordings) and model simulation in the best mean similarity parameters. **H:** Simulated r-a model with the best-matching parameters indicated in panel F. (I=1.9, W=6, b=1). (Right) Phase plane diagram for model with best fit parameters, in the $Excitable_{DOWN}$ regime

characterized by an Excitable$_{UP}$ regime: a stable UP state with noise-induced transitions to a transient DOWN state.

Hippocampus is in an Excitable$_{DOWN}$ regime during NREM sleep

Since the burst-like dynamics of SWR is reminiscent of the Excitable$_{DOWN}$ regime of our model, we asked whether these patterns could also be explained by the same principles. InterSWR durations are much longer (mean = 2.0±0.22s) compared to SWR events (mean = 0.06±0.005s),and more variable ($CV_{InterSWR}$ = 1.3±0.10; CV_{SWR} = 0.33±0.04) (Figure 5E). We applied the duration distribution matching procedure to the SWR/inter-SWR duration distributions and confirmed that the r-a model can also mimic SWR dynamics, with a band of high data-model similarity in the Excitable$_{DOWN}$ regime (Figure 8G). Interestingly, our idealized model is not able to capture the short-interval inter-SWR periods associated with occasional SWR "bursts" (Figure S6H, cite), which suggest the presence of separate SWR-burst promoting mechanisms, possibly arising from interactions with the entorhinal cortex or spatially travelling patterns of SWRs in the hippocampus (Davidson et al., 2009; Yamamoto and Tonegawa, 2017). Accordingly, while the mean ratio and CV_{SWR} of the best fitting model regime were within 2.5 standard deviations of those observed in vivo, the CV of inter-SWR periods was larger than expected from the model. This finding suggests that during NREM sleep the hippocampus is in a tonic DOWN-like state, from which internal 'noise' or an external perturbation can induce spontaneous population-wide spiking events.

Changes in neocortical state correspond to changes in UP state stability

For our initial analysis of the neocortical NREM data we assumed that cortical state is stationary, i.e. absent of variation in model parameters over the course of a single sleep session. However, we found that the duration of UP states showed weak but not insignificant correlation with temporally adjacent UP states, suggestive of a slow process by which the duration of UP states varied over time (Figure S6). EEG recordings from humans during sleep show systematic variation in slow wave dynamics, which has been classified into sleep-depth substates (i.e. stages N1, N2, and N3 or SWS). Rodent sleep is more fragmented, characterized by "packets" of NREM activity (Watson et al., 2016), but has also been classified on a spectrum from deep to light NREM, with power in the LFP delta band (1-4Hz) as a metric for NREM depth, (higher delta power indicating deeper sleep) (Gervasoni et al., 2004). To determine how NREM depth relates to UP/DOWN state durations, we calculated the level of delta power in the 8s time

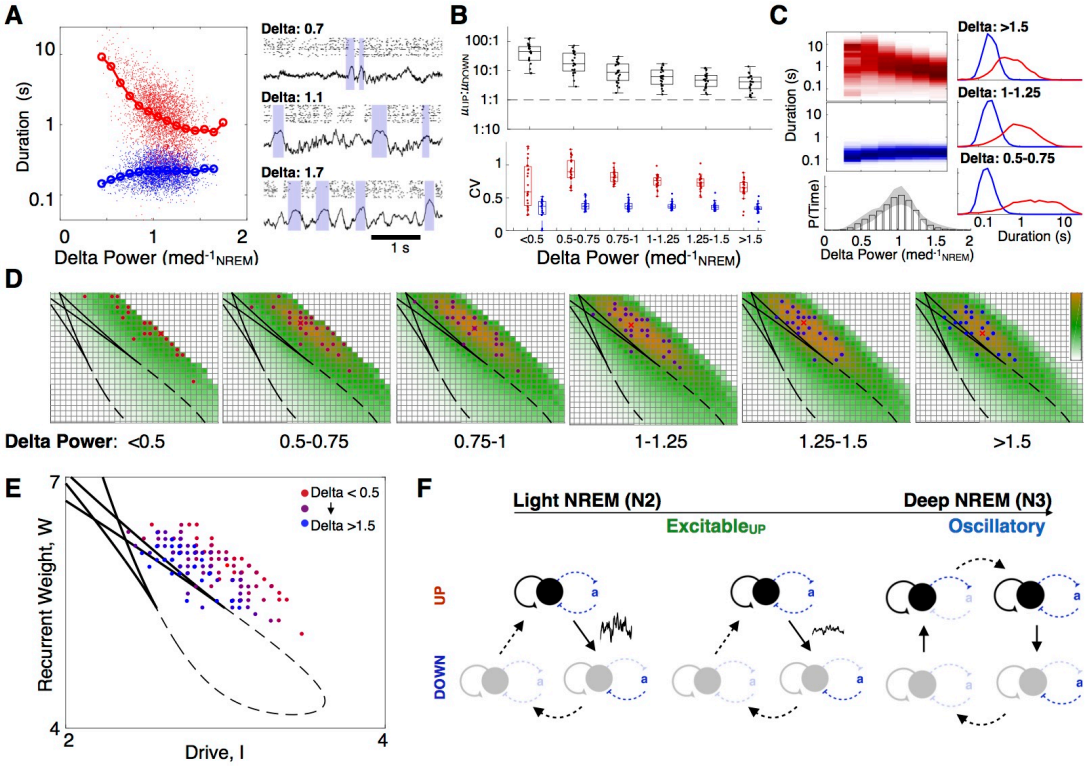

Figure 6 Slow fluctuations in synchronized dynamics during NREM sleep. **A:** UP/DOWN state durations as a function of Delta power in a representative recording. **B:** Ratio of mean UP/DOWN durations in each neocortical recording as a function of delta power (Top). CV of UP/DOWN durations in each neocortical recording as a function of delta power (Bottom). **C:** (Top) Mean distribution of UP and DOWN state durations plotted vs delta power measured in the 8s window around each UP or DOWN state. (Bottom) Mean delta power distribution over all recordings, (error bar: 1 standard deviation). (Right) Mean UP/DOWN state duration distributions for high, medium, and low delta power. **D**: Map of in vivo-model similarity (as in Figure 5B, with τ_r fixed at τ_r=5ms), for the delta-power groups as in panel C. **E:** Best matching parameter value for dwell time distributions in each recording, grouped by delta power (6 groups). Fitting procedure similar to Figure 5, with τ_r fixed at τ_r=5. **F:** Schematic: deeper NREM sleep associated with more stable UP state.

window surrounding each UP and DOWN state (Figure 6A). We found that UP state durations varied to a much greater extent than DOWN state durations with NREM depth (Figure 6A,B,C, S6). While UP state durations were consistently more variable than DOWN state durations regardless of delta power, epochs of lower delta power (light NREM) contained longer UP states, and epochs of higher delta power (deep NREM) were associated with shorter UP states (Figure 6A,B,C, S6).

Indeed, one expects that that delta-band power reflects the incidence of DOWN states, given the relatively consistent DOWN state durations of 100-300ms. As neocortical DOWN states are coincident with slow waves in the EEG, these observations suggest an analogy to the stages of NREM sleep observed in humans, in which lighter stages of NREM (N2) are associated with less frequent slow waves, and deeper stages of sleep (N3/SWS) with more frequent slow waves. In further support of this idea, epochs with intermediate levels of delta power were associated with a heavy tail of infrequent high spindle-power (Figure S6), indicative of the thalamocortical spindle events that are a signature of N2 sleep.

To understand the relationship between NREM depth and UP/DOWN dynamics as captured by our model, we grouped the UP/DOWN states by delta power and calculated data-model similarity maps for UP/DOWN state durations in each group (Figure 6C,D, S7). We found that deeper NREM sleep (higher delta power) was associated with a model parameter domain that is closer to the transition from the Excitable$_{UP}$ to the Oscillatory regime, and that the epochs of highest delta power crossed the boundary into oscillatory dynamics in some sessions. Epochs with the lowest delta power (less than half the median power) were not well fit by the model. The change in parameter domain with deepening sleep corresponded to decreased recurrent excitation, decreased excitability (drive), or increased adaptation strength (Figure S7) in the model. Thus, we find that N2-like sleep corresponds to an Excitable$_{UP}$ regime, and that UP state stability continuously varies with sleep depth resulting in more or less frequent DOWN states. Interestingly, the vast majority of time in all recording sessions was spent in an N2-like Exctitable$_{UP}$ regime (Figure 6B, bottom). Only for a small portion of time in a subset of sessions did cortical state enter an N3-like Oscillatory regime.

Effects of balanced excitation and inhibition on UP/DOWN transitions

Our two-variable model captured significant features of UP/DOWN alternations and the relative roles of adaptation and recurrent excitation. However, firing rate in the UP state was

described only as "active", limited by a saturating input-output function $R_\infty(input)$. On the other hand, neuronal spike rate during the UP state is generally low (Watson et al., 2016) and synaptic inputs during neocortical UP states are balanced between excitation and inhibition (Haider et al., 2006). To understand the contribution of inhibition to UP/DOWN dynamics, we next included an inhibitory population ($\tau_i \approx \tau_e$) into our model (Figure 7A):

$$\tau_e \frac{dr_e}{dt} = -r_e + R_{e,\infty}(w_{ee}r_e - w_{ei}r_i - ba + I_e + \xi_e(t))$$

$$\tau_i \frac{dr_i}{dt} = -r_i + R_{i,\infty}(w_{ie}r_e - w_{ii}r_i + I_i + \xi_i(t))$$

$$\tau_a \frac{da}{dt} = -a + A_\infty(r_e)$$

where adaptation acts on the excitatory population and $R_{e,\infty}(input)$ and $R_{i,\infty}(input)$ are threshold power law I/O relations, as seen in the in vivo-like fluctuation-driven regime (Methods, (Ahmadian et al., 2013)). We find that by incorporating an inhibitory subpopulation in the model we achieve an inhibition-stabilized UP state with low firing rates, and that E-I interaction plays a role in UP/DOWN transitions in this model.

Given that adaptation is slow we can treat a as frozen and visualize model dynamics in the r_e-r_i phase plane (Figure 7B). The fixed point value of r_e as a function of drive describes the effective I/O curve of the excitatory/inhibitory network, r_{ss} (Figure 7C). Like the excitation-only model, recurrent excitation induces bistability at low levels of drive for which the unconnected population would show minimal activation (S7, (Jercog et al., 2017)). In the bistable condition, the r_e-r_i phase plane shows a balanced UP state and a DOWN state fixed point, both stable and separated by a saddle point (Figure 7B,C). Unlike the excitation-only model, the UP state can be stabilized at a low rate determined by the strength of feedback inhibition.

As adaptation effectively decreases the net input to the excitatory population, a plot of r_{ss} vs $I - a$ (Figure 7D), and is analogous to the r-nullcline (rotated 90 degrees) in Figure 5D. Similarly, the curve $A_\infty(r_e)$ vs r_e acts as the analog of the a-nullcline in Figure 5D. Depending on the relative configuration of these two curves, the inhibition-stabilized/adapting model can show UP/DOWN alternations with the same regimes as the two-variable model described above (Figure S7D).

To understand the effect of inhibition in the neocortical slow oscillation, we next investigated the effect of the inhibitory population during Excitable$_{UP}$ dynamics in the adapting, inhibition-stabilized model (Figure 7E). Consider a sufficient perturbation that initiates a transition away from the UP state (Figure 7F). With a dynamic, not frozen, the system is

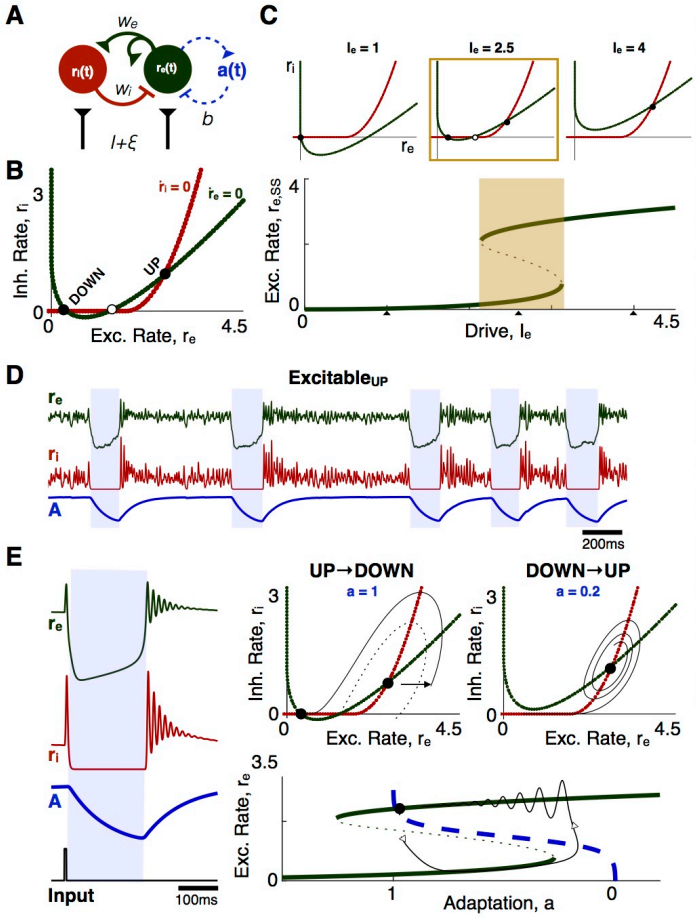

Figure 7 UP/DOWN dynamics in an Adapting Inhibition-Stabilized Network (aISN). **A:** Model schematic. **B:** r_e-r_i phase plane with *a* frozen. **C:** Bistability in the ISN with *a* frozen. Effective I/O curve of the excitatory population rate, $r_{e,ss}$, with r_e-r_i phase planes at low, intermediate, and high levels of drive to the E population (Top). Parameters: w_e = 4, w_i = 3. **D:** Excitable$_{UP}$ dynamics in the aISN. Parameters: $\tau_e = \tau_i$ = 5ms, τ_a= 200ms **E:** UP/DOWN transitions in the Excitable$_{UP}$ ISN. (Left) Time course of the aISN in response to brief excitatory input to the excitatory population. (Bottom Right) Trajectory of the time course in the r_e-*a* phase plane after stimulus-induced transition. The effective I/O curve, $r_{e,ss}$, and the steady state adaptation curve, $A_\infty(r)$, act like the r and a nullclines in the excitation only model. (Top Right) Trajectories of the stimulus-induced UP->DOWN and subsequent DOWN->UP transition in the r_e-r_i phase plane. The dotted line separates the basin of attraction for the DOWN and UP state fixed points.

attracted to the DOWN branch and then drifts along the DOWN branch as adaptation slowly de-activates. When the DOWN state loses stability, the trajectory reaches and rounds the lower knee and transitions abruptly to the only remaining stable solution, the UP branch. Adaptation then builds, returning the system to the low firing rate, E-I balanced UP state (Supplemental video 3).

This E-I balanced Excitable$_{UP}$ regime has implications for the nature of UP->DOWN and DOWN->UP transitions. For our chosen parameter values, small perturbations around the UP state fixed point will exhibit damped, resonant, E-I oscillations as the system returns to the balanced state. The damped oscillations are a result of transient imbalance of excitation and inhibition, and occur when the UP state fixed point is an attracting spiral. As a result, high frequency oscillations occur at the DOWN->UP transition, at a time scale set by the inhibitory population. A further implication is that, counter-intuitively, an UP->DOWN transition can be induced by a sufficiently-strong, brief *excitatory* input to the excitatory population, which can then recruit sufficient inhibition to force the entire network into a DOWN state. This threshold effect is seen in the phase plane as a separatrix: the trajectory shaped as an inverted-U that emerges (in reverse time) from the saddle and curves around the UP state fixed point. It separates the basins of attraction of the UP and DOWN state. From this visualization we see that brief excitatory input to either population can push the trajectory out of the UP state basin of attraction (Figure 7F). Thus, the system can transition to a DOWN state (i.e. a neocortical slow wave) in response to an excitatory perturbation to either population, as well as due to drops in the excitatory population rate.

Discussion

To study cortical dynamics during NREM sleep, we used a firing rate model that represents a general neuronal system with fast positive feedback (recurrent excitation) and slow negative feedback (adaptation). While the model is decidedly idealized, it is amenable to mathematical treatment in terms of a few key parameters, which allows us to develop intuitions for the repertoire of dynamics available to an adapting, recurrent excitatory population. Our analysis of the model revealed an intuitive spectrum of dynamical regimes with UP/DOWN alternations, defined by the stability or transience of UP and DOWN states. Within this dynamical spectrum, we found that the neocortical and hippocampal alternations during NREM sleep are well matched by the model in excitable regimes of dynamics that produce characteristically asymmetric distributions of UP and DOWN state durations (Figure 8A). We next discuss biological interpretations of the model and implications of the findings for UP/DOWN alternations during NREM and other physiological contexts.

Physiological interpretation of model parameters

Our model describes the mean activity of a neuronal population with positive feedback from recurrent excitation (wr), slow negative feedback from adaptation (ba), and a source of noisy drive ($I + \xi(t)$). We interpret drive in the model as the combination of external input and various other factors that drive cells toward spiking, including internal (for example, miniEPSPs) and modulatory influences (for example, increasing the excitability of cells would correspond to an increase in drive parameter in our model). The recurrence parameter, w, reflects the effective weight of excitatory synapses within the population. This self-excitation drives the population during the UP state. In the two-variable model, the saturating I/O relation, $R_\infty(input)$, imposes a maximum activity at $r(t) = 1$. However, the three-variable model reveals that fast negative feedback from inhibitory cells can dynamically stabilize the rate of excitatory population during the UP state.

Adaptation could encompass a variety of physiological processes that are activated by, and subsequently reduce, spiking at a slow timescale (~50ms-1s), such as slow voltage or calcium activated potassium currents, or sodium current inactivation (Jung et al., 1997; Vergara et al., 1998). For mathematical simplicity, we have reduced the effect of adaptive processes, to a single, saturating, mean field variable, $a(t)$, that negatively feeds back on population activity with strength b. Previous studies modeled synaptic or slow divisive feedback to achieve the same goal (Holcman and Tsodyks, 2006; Tabak et al., 2011), in some cases showing

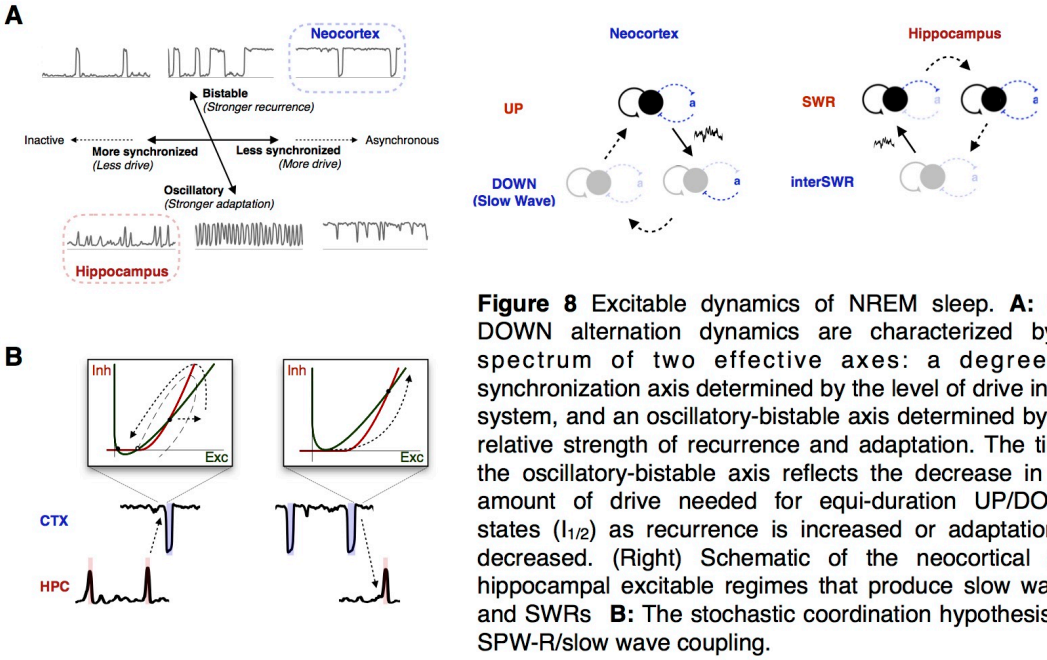

Figure 8 Excitable dynamics of NREM sleep. **A:** UP/DOWN alternation dynamics are characterized by a spectrum of two effective axes: a degree of synchronization axis determined by the level of drive in the system, and an oscillatory-bistable axis determined by the relative strength of recurrence and adaptation. The tilt of the oscillatory-bistable axis reflects the decrease in the amount of drive needed for equi-duration UP/DOWN states ($I_{1/2}$) as recurrence is increased or adaptation is decreased. (Right) Schematic of the neocortical and hippocampal excitable regimes that produce slow waves and SWRs **B:** The stochastic coordination hypothesis for SPW-R/slow wave coupling.

characteristic differences (Tabak et al., 2011). Further work to identify signatures of different sources of adaptive feedback could give insight to the implications and identification of biophysical substrates of adaptation in different physiological contexts.

With these idealizations, our model encompasses numerous previous models for UP/DOWN dynamics, from mean field to large scale spiking models. We have reduced the critical influences to a few key parameters that provide an intuitive understanding of a wide range of UP/DOWN alternation dynamics in neuronal populations.

General insight to synchronized UP/DOWN dynamics

Due to their ubiquity under conditions of low neuromodulatory tone, UP/DOWN alternation dynamics are proposed to represent a "default state" of cortical tissue (Buzsáki, 2006; Sanchez-Vives et al., 2017), a view supported by their presence in isolated cortical and hippocampal tissue (Buzsáki et al., 1987; Timofeev et al., 2000). Our model reveals why UP/DOWN alternations are so ubiquitous: they are seen in neural populations with sufficiently strong local recurrent excitation (>w_0) at comparatively low levels of drive ($\sim I_{1/2}$).

Synchronized UP/DOWN dynamics have been proposed to exist in a "multi-dimensional spectrum" (Harris and Thiele, 2011). Our model captures a subspace of this spectrum defined by two axes (Figure 8A). First, the magnitude of excitatory drive determines the degree of synchronization (Mochol et al., 2015). Increasing drive brings the population from a DOWN-dominated "more synchronized" regime with brief population bursts, to an UP-dominated "more asynchronous" regime with occasional DOWN states. The relative strength of recurrent excitation and adaptation determines the temporal dynamics between these extremes. When recurrence dominates, the system is characterized by bistability; when adaptation dominates, oscillations emerge. We note that this second axis also influences the "steepness" of UP/DOWN transitions, even in the adjacent excitable regimes, which we have not explored.

Different experimental conditions are associated with different regimes in the spectrum of UP/DOWN dynamics captured by our model. Other models have described oscillatory UP/DOWN dynamics in the spinal cord (Vladimirski et al., 2008) and in cortical slices (Mattia and Sanchez-Vives, 2012), whereas in cortical cultures $\text{Excitable}_{\text{DOWN}}$-like bursting behavior dominates (Hinard et al., 2012). Recently, bistable UP/DOWN dynamics were found to describe the activity of sensory cortex during urethane anesthesia (Jercog et al., 2017; Mochol et al., 2015). During quiet wakefulness, cortical state varies following the level of arousal of the animal

(McGinley et al., 2015), but is unclear which of the regimes. Our model provides a framework by which one can predict how duration statistics should change with experimental manipulation of intrinsic or network properties, for example by varying levels of anesthesia or applying other pharmacological agents.

Extensive study has revealed multiple factors that can bring cortical tissue from the "default mode" of UP/DOWN dynamics to an activated state. In culture and slice preparations, increased levels of subcortical neuromodulators are able to 'wake up' the tissue, leading to the replacement of slow oscillations in the neocortex with asynchronous spiking (Hinard et al., 2012) and SWRs in the hippocampus with theta-like oscillations (Konopacki et al., 1987). Acetylcholine induces neocortical desynchronization and hippocampal theta during REM sleep ((Jouvet, 1994)), and acetylcholine, norepinephrine, or thalamic drive can give rise to cortical desynchronization during quiet wakefulness (Eggermann et al., 2014; Poulet and Petersen, 2008). Each of these factors have effects that correspond to parameter changes that can transition our model from UP/DOWN dynamics to an asynchronous (tonic UP) state: 1) increasing excitability or external drive, I, 2) decreasing recurrence below the critical value for UP/DOWN alternations, $w < w_0$, or 3) decreasing the strength of adaptation, b, or 4) increasing recurrence such that the DOWN state loses stability. Similarly, the ascending neuromodulators increase excitability by depolarizing cortical pyramidal cells, decrease the effective weight of excitatory-excitatory synapses, and deactivate adaptive currents (Cole and Nicoll, 1983; Hasselmo and McGaughy, 2004). By these parallels, we are able to apply our model to interpret the putative mechanisms by which cortical desynchronization follows global and local levels of arousal.

UP/DOWN dynamics of the NREM slow oscillation

Our recordings from naturally-sleeping rats show that the neocortical UP/DOWN dynamics during NREM are unlike those reported during anesthesia or quiet wakefulness. Despite the widely used term slow "oscillation" (Steriade et al., 1993), we found that the NREM slow oscillation is reflective of $\text{Excitable}_{\text{UP}}$ dynamics: an irregular process in which activity fluctuations during stable UP states can lead to transient DOWN states. Furthermore, we found that the stability of the UP state is not constant but evolves over the course of sleep in a manner that resembles the stages of NREM/SWS sleep in humans (Berry et al., 2017). In relatively superficial NREM sleep (N1 stage, using the human clinical term), long UP states are occasionally punctuated by neuronal silence-associated positive delta or "slow waves" (surface

negative, depth positive), which can be localized at one or few recording sites (Sirota and Buzsáki, 2005). As sleep deepens, the incidence of DOWN states increase and they become synchronous over larger cortical areas (N2 stage; (Nir et al., 2011)) The DOWN-UP transitions occasionally become strongly synchronous, producing a sharp LFP wave (surface positive, depth negative), known as the K complex (Cash et al., 2009). With further deepening of sleep, DOWN states become more frequent and short episodes of repeating DOWN states may become rhythmic (N3 stage). This relationship explains the inverse correlation between delta power, measuring mainly the large LFP deflections of the DOWN state, and UP state duration. By our model, the decreasing stability of the UP state with deeper stages of sleep may be due to 1) decreased recurrent strength, 2) increased adaptation, or 3) decreased excitability. Quantifying the time spent in these sub-states (N1, N2, N3) revealed that the N3 state in the rat occupies only a small fraction of NREM sleep, whereas in humans this stage is more prominent. Overall, our model captures the entire qualitative dynamic of NREM evolution. However, the quantitative share of the various NREM stages differ between rodent and human sleep.

While the model is ambiguous to the biophysical substrate of adaptation, we can make some predictions: first, the adaptive process responsible for neocortical UP/DOWN alternations should be constitutively active during the UP state and deactivate during the hyperpolarized DOWN state. Because of low spike rates during the NREM UP state, subthreshold adaptation conductances are a good candidate, especially given that most neurons are silent or fire at a very low rate during any given UP state. Adaptation in our model could also include effects of a hyperpolarization-activated excitatory process, such as I_h. Second, the neocortical adaptive process should recover at a time scale reflective of the DOWN state duration (~200ms). Our model predicts how changing properties of the relevant adaptive mechanisms experimentally or in disease should change the duration statistics of the slow oscillation, which we hope can provide a guide for future experiments to uncover the biophysical substrates of these physiologically relevant dynamics.

Regional differences that support differential neocortical and hippocampal dynamics

The different mean durations of the neocortical DOWN state and the hippocampal SWR indicate that alternation dynamics is mostly likely mediated by different adaptive processes in the two regions. The hippocampal adaptive process should activate at a time scale that reflects the SWR durations (~60ms), faster than the one that mediates the neocortical slow oscillation. Previous work has revealed threshold behavior in the generation of SWRs, indicative of

Excitable$_{DOWN}$ dynamics, with a GABA$_B$-mediated adaptation mechanism (English et al., 2014; Menendez de la Prida et al., 2006).

An important question is how neocortical and hippocampal networks would occupy different dynamic regimes of the same model. Neocortical slow oscillations and hippocampal SWRs are present simultaneously during NREM sleep. Although they appear fundamentally different, our model with just a few parameters can quantitatively account for both by using different parameter values. We suggest that the different nature of recurrent connectivity in the two regions is responsible for their differing dynamics.

The neocortex is a modularly organized structure. In contrast, the hippocampus can be conceived as a single expanded cortical module (Amaral and Witter, 1989). Strongly recurrent pyramidal cell populations are found in neocortical layer 5 and the hippocampal CA2 and CA3a subregions (Li et al., 1994), which would support their role as the locus of UP state and sharp wave initiation, respectively (Buzsáki et al., 1983; Sanchez-Vives and McCormick, 2000). However, crucial differences exist between connectivity of neocortical layer 5 and hippocampal CA2-3 regions. Excitatory connectivity in layer 5 is local (200 μm), dense (up to 80% connection probability), and follows a 'Mexican hat' excitatory-inhibitory spatial structure with strong local excitatory connections and spatially extensive inhibition (Markram et al., 2004). In contrast, excitatory connectivity in the hippocampus is sparse and spatially extensive(Li et al., 1994), with local inhibitory connections ((English et al., 2017; Freund and Buzsáki, 1998)). While layer 5 excitatory synapses are relatively strong, the transmitter release probability of synapses between hippocampal pyramidal neurons is very low, resulting in comparatively weak synapses (Silver, 2010). Together, these factors indicate that the effective strength of recurrence in the hippocampus is lower than that in neocortex, which would result in DOWN-dominated as opposed to UP-dominated dynamics, as are observed. To demonstrate that connectivity and synaptic strength are responsible for the different NREM dynamics of the neocortex and hippocampus will require experimental manipulations that explicitly vary these parameters in combination with models that take into account the spatial effects of recurrent excitation.

NREM function through stochastic coordination of excitable dynamics

According to the two-stage model of memory consolidation (Buzsáki, 1989; McClelland et al., 1995), the hippocampus acts as a fast, but unstable, learning system. In contrast, the neocortex acts as a slow learning system that forms long-lasting memories after many presentations of a stimulus. The two-stage model proposes that recently-learned patterns of

activity are reactivated in the hippocampus during SWRs, which act as a "training" signal for the neocortex, and that the neocortical consolidation of those patterns relies on SWR-slow wave coupling (Maingret et al., 2016; Peyrache et al., 2009). The excitable dynamics described provide a mechanism by which the stochastic, or probabilistic, coordination of state transitions in each structure could support communication between the two regions (Figure 8B). Namely, the excitatory kick of a hippocampal SWR could induce a neocortical UP->DOWN transition by briefly disrupting the neocortical excitatory/inhibitory balance.

Extensive experimental evidence points towards temporal coordination between slow waves and SWRs. Slow waves in higher-order neocortical regions are more likely following SWRs (Battaglia et al., 2004; Peyrache et al., 2009), and SWR->slow wave coupling is associated with reactivation of recently learned activity patterns in the neocortex (Ji and M. A. Wilson, 2007; Maingret et al., 2016; Peyrache et al., 2009). Interestingly, slow waves are not reliably evoked by SWRs, suggesting that the efficacy of any one SWR to evoke a slow wave in a given cortical region is probabilistic. The probability of SWR->slow wave induction likely varies by brain state, cortical region, and even SWR spiking content. Further work to investigate how these factors shape SWR->slow wave coupling will likely shed light on the brain-wide mechanisms of memory consolidation.

How then, does a SWR-induced neocortical slow wave induce changes in the neocortex? Recent work suggests that the DOWN->UP transition at the end of a slow wave (aka the k complex) is a window of opportunity for synaptic plasticity that supports NREM functions (Levenstein et al., 2017). Spike sequences during the DOWN->UP transition produce lasting changes in synaptic strength, that correlate with post-sleep behavioral change (Gulati et al., 2017; Kruskal et al., 2013; Wei et al., 2016). SWR-coupled slow waves have altered spiking dynamics at the subsequent DOWN->UP transition (Maingret et al., 2016). Interestingly, we found in our model that the same E-I balance that allows excitatory input such as a SWR to initiate a neocortical UP->DOWN transition produces a burst of resonant excitatory-inhibitory (gamma-like) oscillation at the DOWN->UP transition. High gamma (~60-150Hz) activity is seen following slow waves *in vivo* (Watson et al., 2016) and may act to coordinate and promote plasticity between cell assemblies (Buzsáki and Wang, 2012).

In turn, the burst of activity at the neocortical DOWN->UP transition could induce SWRs in the hippocampus. The functional role of slow wave->SWR coupling is less well understood, but hippocampal SWRs are more likely immediately following slow waves in some neocortical regions - including the entorhinal cortex (Isomura et al., 2006; Peyrache et al., 2009). Slow

wave->SWR could provide a mechanism by which neocortical activity is able to bias SWR content, or another mechanism by which the SWR could interact with the neocortical window of plasticity at the DOWN->UP transition. Further, a SWR-slow wave-SWR loop could produce the occasional SWR bursts not captured by our model of hippocampal SWR activity in isolation. Uncovering regional or state-dependent differences in the directionality of coupling could provide insight into the mechanisms that support of memory consolidation.

Conclusions

Our results reveal that NREM sleep is characterized by structure-specific excitable dynamics in the mammalian forebrain. We found that a model of an adapting recurrent neural population is sufficient to capture a variety of UP/DOWN alternation dynamics comparable to those observed *in vivo*. The neocortical "slow oscillation" is well-matched by the model in an $\text{Excitable}_{\text{UP}}$ regime in which a stable UP state is punctuated by transient DOWN states, while the hippocampal sharp waves are well-matched by the model in an $\text{Excitable}_{\text{DOWN}}$ regime in which a stable DOWN state is punctuated by transient UP states (Figure 8A). With balanced excitation and inhibition, UP->DOWN transitions can be induced via excitatory perturbations that momentarily disrupt balance, and DOWN->UP transitions are followed by a high frequency oscillation as the system returns to the balanced state. Furthermore, we found that slow fluctuations in neocortical state over the course of sleep correspond to slow fluctuations in the stability of the neocortical UP state. Our results offer a unifying picture of hippocampal and neocortical dynamics during NREM sleep, and suggest that NREM function is mediated by excitable dynamics in the two structures.

Acknowledgements

The authors would like to thank Rachel Swanson, William Muñoz, Brendon Watson, Andres Grosmark for discussions during the development of the project and extensive feedback on the manuscript, the NIH training grant for computational neuroscience T90DA043219 for funding and the trainees for their feedback on the manuscript, and Brendon Watson and Andres Grosmark for generously sharing their data.

Author Contributions

Conceptualization, DL, GB, JR; Methodology, DL JR; Formal Analysis, DL JR; Writing – Original Draft, DL, JR; Writing – Review & Editing, DL, GB, JR; Supervision, GB and JR.

Declaration of Interests

The authors declare no competing interests.

REFERENCES

Achermann, P., Borbely, A.A., 1997. Low-frequency (< 1 Hz) oscillations in the human sleep electroencephalogram. Neuroscience

Ahmadian, Y., Rubin, D.B., Miller, K.D., 2013. Analysis of the stabilized supralinear network. Neural Computation

Amaral, D.G., Witter, M.P., 1989. The three-dimensional organization of the hippocampal formation: A review of anatomical data. Neuroscience

Ball, G.J., Gloor, P., Schaul, N., 1977. The cortical electromicrophysiology of pathological delta waves in the electroencephalogram of cats. Electroencephalogr Clin Neurophysiol

Battaglia, F.P., Sutherland, G.R., McNaughton, B.L., 2004. Hippocampal sharp wave bursts coincide with neocortical “up-state” transitions. Learn. Mem

Bazhenov, M., Timofeev, I., Steriade, M., Sejnowski, T., 2002. Model of thalamocortical slow-wave sleep oscillations and transitions to activated States. Journal of Neuroscience

Berry, R.B., Brooks, R., Gamaldo, C., Harding, S.M., Lloyd, R.M., Quan, S.F., Troester, M.T., Vaughn, B.V., 2017. AASM Scoring Manual Updates for 2017 (Version 2.4). J Clin Sleep Med

Buzsáki, G., 2015. Hippocampal sharp wave-ripple: A cognitive biomarker for episodic memory and planning. Hippocampus

Buzsáki, G., 2006. Rhythms of the Brain.

Buzsáki, G., 1989. Two-stage model of memory trace formation: a role for “noisy” brain states. Neuroscience.

Buzsáki, G., Bickford, R.G., Ponomareff, G., Thal, L.J., Mandel, R., Gage, F.H., 1988. Nucleus basalis and thalamic control of neocortical activity in the freely moving rat. The Journal of Neuroscience

Buzsáki, G., Czopf, J., KondÀkor, I., Björklund, A., Gage, F.H., 1987. Cellular activity of intracerebrally transplanted fetal hippocampus during behavior. Neuroscience

Buzsáki, G., Leung, L.W., Vanderwolf, C.H., 1983. Cellular bases of hippocampal EEG in the behaving rat. Brain Research.

Buzsáki, G., Wang, X.J., 2012. Mechanisms of gamma oscillations. Annu. Rev. Neurosci.

Cash, S.S., Halgren, E., Dehghani, N., Rossetti, A.O., Thesen, T., Wang, C., Devinsky, O., Kuzniecky, R., Doyle, W., Madsen, J.R., Bromfield, E., Erőss, L., Halász, P., Karmos, G., Csercsa, R., Wittner, L., Ulbert, I., 2009. The human K-complex represents an isolated cortical down-state. Science

Cole, A.E., Nicoll, R.A., 1983. Acetylcholine mediates a slow synaptic potential in hippocampal pyramidal cells. Science

Colgin, L.L., Kubota, D., Jia, Y., Rex, C.S., Lynch, G., 2004. Long-term potentiation is impaired in rat hippocampal slices that produce spontaneous sharp waves. The Journal of Physiology

Compte, A., Sanchez-Vives, M.V., McCormick, D.A., Wang, X.J., 2003. Cellular and network mechanisms of slow oscillatory activity (<1 Hz) and wave propagations in a cortical network model. Journal of Neurophysiology

Contreras, D., Timofeev, I., Steriade, M., 1996. Mechanisms of long-lasting hyperpolarizations underlying slow sleep oscillations in cat corticothalamic networks. The Journal of Physiology

Curto, C., Sakata, S., Marguet, S., Itskov, V., Harris, K.D., 2009. A simple model of cortical dynamics explains variability and state dependence of sensory responses in urethane-anesthetized auditory cortex. Journal of Neuroscience

Davidson, T.J., Kloosterman, F., Wilson, M.A., 2009. Hippocampal replay of extended experience. Neuron

Destexhe, A., 2009. Self-sustained asynchronous irregular states and Up-Down states in

thalamic, cortical and thalamocortical networks of nonlinear integrate-and-fire neurons. J Comput Neurosci

Eggermann, E., Kremer, Y., Crochet, S., Petersen, C.C., 2014. Cholinergic Signals in Mouse Barrel Cortex during Active Whisker Sensing. Cell Rep

English, D.F., McKenzie, S., Evans, T., Kim, K., Yoon, E., Buzsáki, G., 2017. Pyramidal Cell-Interneuron Circuit Architecture and Dynamics in Hippocampal Networks. Neuron

English, D.F., Peyrache, A., Stark, E., Roux, L., Vallentin, D., Long, M.A., Buzsáki, G., 2014. Excitation and Inhibition Compete to Control Spiking during Hippocampal Ripples: Intracellular Study in Behaving Mice. Journal of Neuroscience 34, 16509–16517. doi:10.1523/JNEUROSCI.2600-14.2014

Freund, T.F., Buzsáki, G., 1998. Interneurons of the hippocampus. Hippocampus

Gervasoni, D., Lin, S.-C., Ribeiro, S., Soares, E.S., Pantoja, J., Nicolelis, M.A.L., 2004. Global forebrain dynamics predict rat behavioral states and their transitions. Journal of Neuroscience

Grosmark, A.D., Buzsáki, G., 2016. Diversity in neural firing dynamics supports both rigid and learned hippocampal sequences. Science

Gulati, T., Guo, L., Ramanathan, D.S., Bodepudi, A., Ganguly, K., 2017. Neural reactivations during sleep determine network credit assignment. Nature Publishing Group

Haider, B., Duque, A., Hasenstaub, A.R., McCormick, D.A., 2006. Neocortical network activity in vivo is generated through a dynamic balance of excitation and inhibition. Journal of Neuroscience

Harris, K.D., Thiele, A., 2011. Cortical state and attention. Nat Rev Neurosci

Hasselmo, M.E., McGaughy, J., 2004. High acetylcholine levels set circuit dynamics for attention and encoding and low acetylcholine levels set dynamics for consolidation. Prog. Brain Res.

Hill, S., Tononi, G., 2005. Modeling sleep and wakefulness in the thalamocortical system. Journal of Neurophysiology

Hinard, V., Mikhail, C., Pradervand, S., Curie, T., Houtkooper, R.H., Auwerx, J., Franken, P., Tafti, M., 2012. Key electrophysiological, molecular, and metabolic signatures of sleep and wakefulness revealed in primary cortical cultures. Journal of Neuroscience

Holcman, D., Tsodyks, M.V., 2006. The emergence of Up and Down states in cortical networks. PLoS Comp Biol

Isomura, Y., Sirota, A., Ozen, S., Montgomery, S., Mizuseki, K., Henze, D.A., Buzsáki, G., 2006. Integration and segregation of activity in entorhinal-hippocampal subregions by neocortical slow oscillations. Neuron

Jercog, D., Roxin, A., Barthó, P., Luczak, A., Compte, A., la Rocha, de, J., 2017. UP-DOWN cortical dynamics reflect state transitions in a bistable network. eLife.

Ji, D., Wilson, M.A., 2007. Coordinated memory replay in the visual cortex and hippocampus during sleep.

Jouvet, M., 1994. The Paradox of Sleep: The Story of Dreaming. Translated by Laurence Garey (1999).

Jung, H.Y., Mickus, T., Spruston, N., 1997. Prolonged sodium channel inactivation contributes to dendritic action potential attenuation in hippocampal pyramidal neurons. The Journal of Neuroscience

Konopacki, J., Maciver, M.B., Bland, B.H., Roth, S.H., 1987. Theta in hippocampal slices: relation to synaptic responses of dentate neurons. Brain Research Bulletin

Kruskal, P.B., Li, L., MacLean, J., 2013. Circuit reactivation dynamically regulates synaptic plasticity in neocortex. Nature Communications 4. doi:10.1038/ncomms3574

Latham, P.E., Richmond, B.J., Nelson, P.G., Nirenberg, S., 2000. Intrinsic dynamics in neuronal

networks. I. Theory. Journal of Neurophysiology
Lee, S.-H., Dan, Y., 2012. Neuromodulation of brain states. Neuron
Levenstein, D., Watson, B.O., Rinzel, J., Buzsáki, G., 2017. Sleep regulation of the distribution of cortical firing rates. Curr Opin Neurobiol 44, 34–42. doi:10.1016/j.conb.2017.02.013
Li, X.G., Somogyi, P., Ylinen, A., Buzsáki, G., 1994. The hippocampal CA3 network: an in vivo intracellular labeling study. J. Comp. Neurol.
Maingret, N., Girardeau, G.J., Todorova, R., Goutierre, M., Zugaro, M., 2016. Hippocampo-cortical coupling mediates memory consolidation during sleep. Nature neuroscience
Markram, H., Toledo-Rodriguez, M., Wang, Y., 2004. Interneurons of the neocortical inhibitory system. Nature Reviews ….
Mattia, M., Sanchez-Vives, M.V., 2012. Exploring the spectrum of dynamical regimes and timescales in spontaneous cortical activity. Cogn Neurodyn
McClelland, J.L., McNaughton, B.L., O'Reilly, R.C., 1995. Why there are complementary learning systems in the hippocampus and neocortex: insights from the successes and failures of connectionist models of learning and memory. Psychological Review
McGinley, M.J., David, S.V., McCormick, D.A., 2015. Cortical Membrane Potential Signature of Optimal States for Sensory Signal Detection. Neuron
Menendez de la Prida, L., Huberfeld, G., Cohen, I., Miles, R., 2006. Threshold Behavior in the Initiation of Hippocampal Population Bursts. Neuron
Mochol, G., Hermoso-Mendizabal, A., Sakata, S., Harris, K.D., la Rocha, de, J., 2015. Stochastic transitions into silence cause noise correlations in cortical circuits. PNAS
Nir, Y., Staba, R.J., Andrillon, T., Vyazovskiy, V.V., Cirelli, C., Fried, I., Tononi, G., 2011. Regional Slow Waves and Spindles in Human Sleep. Neuron
Parga, N., Abbott, L.F., 2007. Network model of spontaneous activity exhibiting synchronous transitions between up and down States. Frontiers in Neuroscience
Petersen, C.C., Hahn, T.T.G., Mehta, M., Grinvald, A., Sakmann, B., 2003. Interaction of sensory responses with spontaneous depolarization in layer 2/3 barrel cortex. Proc. Natl. Acad. Sci.
Peyrache, A., Khamassi, M., Benchenane, K., Wiener, S.I., Battaglia, F.P., 2009. Replay of rule-learning related neural patterns in the prefrontal cortex during sleep. Nature neuroscience
Poulet, J., Petersen, C.C., 2008. Internal brain state regulates membrane potential synchrony in barrel cortex of behaving mice. Nature
Sanchez-Vives, M.V., Massimini, M., Mattia, M., 2017. Shaping the Default Activity Pattern of the Cortical Network. Neuron
Sanchez-Vives, M.V., McCormick, D.A., 2000. Cellular and network mechanisms of rhythmic recurrent activity in neocortex. Nature neuroscience
Siapas, A.G., Wilson, M.A., 1998. Coordinated Interactions between Hippocampal Ripples and Cortical Spindles during Slow-Wave Sleep. Neuron
Silver, R.A., 2010. Neuronal arithmetic. Nat Rev Neurosci
Sirota, A., Buzsáki, G., 2005. Interaction between neocortical and hippocampal networks via slow oscillations. Thalamus Relat Sys
Sirota, A., Csicsvari, J., Buhl, D., Buzsáki, G., 2003. Communication between neocortex and hippocampus during sleep in rodents. Proc. Natl. Acad. Sci.
Steriade, M., Nunez, A., Amzica, F., 1993. A novel slow (< 1 Hz) oscillation of neocortical neurons in vivo: depolarizing and hyperpolarizing components. The Journal of Neuroscience
Strogatz, S.H., 2014. Nonlinear dynamics and chaos: with applications to physics, biology, chemistry, and engineering.
Tabak, J., Rinzel, J., Bertram, R., 2011. Quantifying the relative contributions of divisive and subtractive feedback to rhythm generation. PLoS Comp Biol

Timofeev, I., Grenier, F., Bazhenov, M., Sejnowski, T., Steriade, M., 2000. Origin of slow cortical oscillations in deafferented cortical slabs. Cereb. Cortex

Vergara, C., Latorre, R., Marrion, N.V., Adelman, J.P., 1998. Calcium-activated potassium channels. Curr Opin Neurobiol

Vladimirski, B.B., Tabak, J., O'Donovan, M.J., Rinzel, J., 2008. Episodic activity in a heterogeneous excitatory network, from spiking neurons to mean field. J Comput Neurosci

Watson, B.O., Ding, M., Buzsáki, G., 2017. Temporal coupling of field potentials and action potentials in the neocortex. European Journal of Neuroscience

Watson, B.O., Levenstein, D., Greene, J.P., Gelinas, J.N., Buzsáki, G., 2016. Network Homeostasis and State Dynamics of Neocortical Sleep. Neuron

Wei, Y., Krishnan, G.P., Bazhenov, M., 2016. Synaptic Mechanisms of Memory Consolidation during Sleep Slow Oscillations. Journal of Neuroscience

Wilson, H.R., Cowan, J.D., 1972. Excitatory and inhibitory interactions in localized populations of model neurons. Biophysical journal

Yamamoto, J., Tonegawa, S., 2017. Direct Medial Entorhinal Cortex Input to Hippocampal CA1 Is Crucial for Extended Quiet Awake Replay. Neuron

Ylinen, A., Bragin, A., Nádasdy, Z., Jandó, G., Szabó, I., Sik, A., Buzsáki, G., 1995. Sharp wave-associated high-frequency oscillation (200 Hz) in the intact hippocampus: network and intracellular mechanisms. The Journal of Neuroscience

METHODS

Datasets

The datasets used were reported in Watson et al 2016 (neocortex) and Grosmark and Buzsaki 2016 (hippocampus), and are briefly summarized here.

For the cortical dataset, silicon probes were implanted in frontal cortical areas of 11 male Long Evans rats. Recording sites included medial prefrontal cortex, anterior cingulate cortex, premotor cortex/M2, and orbitofrontal cortex. Neural activity during natural sleep-wake behavior was recorded using high-density silicon probes during light hours in the animals' home cage. 25 recordings of mean duration 4.8+/-2.2hrs were recorded. The raw 20kHz data was low-pass filtered and resampled at 1250Hz to extract local field potential information. To extract spike times, the raw data high-pass filtering at 800Hz, and then threshold-crossings were detected. KlustaKwik software was used to cluster spike waveforms occurring simultaneously on nearby recording sites, and Klusters software was used for manual inspection of waveforms consistent with a single neuronal source. Units were classified into putative excitatory (pE) and putative inhibitory (pI) based on the spike waveform metrics. Each animal had 35+/-12 detected pE units and 5+/-3 detected pI units on average. For the hippocampal dataset, silicon probes were implanted in the dorsal hippocampus of 4 male Long Evans rats. Neural activity during sleep was recorded before and after behavior on a linear track. LFP and spikes were extracted similar to the cortical dataset. Sharp-wave ripple events were detected as described in Grosmark and Buzsaki 2016.

NREM Detection

Sleep state was detected using an automated scoring algorithm as described previously (Watson et al 2016), with some modifications. As only the NREM state was used in this study, we describe here the process for NREM detection. However, the code for full state detection is available at https://github.com/buzsakilab/buzcode. NREM sleep was detected using the FFT spectrogram of a neocortical LFP channel, calculated in overlapping 10s windows at 1s intervals. Power in each time window was calculated for frequencies that were logarithmically spaced from 1 to 100Hz. The spectral power was then log transformed, and z-scored over time for each frequency. The slow wave power (signature of NREM sleep) was calculated by weighting each frequency by a weight determined from the mean of the weights for the first principal components from the dataset in Watson et al 2016, which was found to distinguish NREM and non-NREM in all recordings. While the same dataset was used here, using the filter

(i.e. weighted frequency)-based approach as opposed to PCA makes the algorithm robust for a wider range of recording conditions, especially those in which there is less time spent asleep (and thus NREM may not be expected to account for the largest portion of variance). Like the first principal component, the slow wave filtered signal was found to be bimodal in all recordings, and the lowest point between modes of the distribution was used to divide NREM and non-NREM epochs.

In the hippocampal dataset, manual NREM scoring as reported in Grosmark and Buzsaki 2016 was used for this study.

Slow Wave Detection

Slow waves were detected using the coincidence of a two-stage threshold crossing in two signals (Figure S1A,B): a drop in high gamma power (100-400Hz, representative of spiking (Watson et al 2017)) and a peak in the delta-band filtered signal (0.5-8Hz). The gamma power signal was smoothed using a sliding 80ms window, and locally normalized using a modified (non-parametric) Z-score in the surrounding 20s window, to account for non-stationaries in the data (for example due to changes in brain state and noise), that could result in local fluctuations in gamma power. The channel used for detection was determined as the channel for which delta was most negatively correlated with spiking activity, while gamma was most positively correlated with spiking activity.

Two thresholds were used for event detection in each LFP-derived signal, a “peak threshold” and a “window threshold”. Time epochs in which the delta-filtered signal crossed the peak threshold were taken as putative slow wave events, with start and end times at the nearest crossing of the window threshold. Peak/window thresholds were determined for each recording individually to best give separation between spiking (UP states) and non-spiking (DOWN states) (Figure S1C). To determine the delta thresholds, all peaks in the delta-filtered signal greater than 0.25 standard deviations were detected as candidate delta peaks and binned by peak magnitude. The peri-event time histogram (PETH) for spikes from all cells was calculated around delta peaks in each magnitude bin, and normalized by the mean rate in all bins. The smallest magnitude bin at which spiking (i.e. the PETH at time = 0) was lower than a set rate threshold (the “sensitivity” parameter, Figure S1D) was taken to be the peak threshold. For example, a sensitivity of 0.5 means that the delta peak threshold is set to the smallest threshold for which spiking drops below 50% of mean spiking activity. The window threshold was set to the average delta value at which the rate crosses this threshold in all peak magnitude bins. The

gamma thresholds were calculated similarly, but using drops below a gamma power magnitude instead of peaks above a delta magnitude.

Once the thresholds were calculated, candidate events were then detected in the delta and gamma power signals, and further limited to a minimum duration of 40ms. Slow wave events were then taken to be overlapping intervals of both the gamma and delta events. DOWN states with spiking above the sensitivity threshold were thrown out.

Detection quality was checked using a random sampling and visual inspection protocol. LFP and spike rasters for random 10s windows of NREM sleep were presented to a manual scorer, who marked correct SW detections, false alarms, and missed SWs. This protocol was used to estimate the detection quality (miss %, FA %) for each recording (Figure S1E), and to optimize the detection algorithm.

Model Equations

The r-a model represents the mean firing rate or activity of a neural population with activity-driven adaptation a(t).

$$\tau_r \dot{r} = -r + R_\infty\big(wr - ba + I + \xi(t)\big) \qquad (1)$$

$$\tau_a \dot{a} = -a + A_\infty(r) \qquad (2)$$

with activation functions

$$R_\infty(x) = \frac{1}{1 + e^{-(x - x_0)}}$$

and

$$A_\infty(r) = \frac{1}{1 + e^{-k(r - r_0)}}$$

Unless otherwise specified, we use $x_0 = 5$, $r_0 = 0.5 = R_\infty(x_0)$, and $k = 15$ to parameterize the activation functions. Values of other parameters are as indicated in the text and figure legends. The rate and adaptation variables can be considered as non-dimensionalized, scaled by their maximum possible values; similarly, we use $\tau_r = 1$ (unless otherwise specified) so that time is dimensionless (AU, arbitrary units), scaled by the time constant for firing rate.

The e-i-a model represents the mean rate of an adapting excitatory and an inhibitory population,

$$\tau_e \dot{r_e} = -r_e + R_{e,\infty}\big(w_{ee} r_e - w_{ei} r_i - ba + I_e + \xi_e(t)\big) \qquad (3)$$

$$\tau_i \dot{r_i} = -r_i + R_{i,\infty}\left(w_{ie} r_e - w_{ii} r_i + I_i + \xi_i(t)\right) \quad (4)$$

$$\tau_a \dot{a} = -a + A_\infty(r_e) \quad (5)$$

with power law activation functions, as in Ahmadian and Milller (2013).

$$R_{e/i,\infty}(x) = k[x]_+^{\ n}$$

The activation function of adaptation, $A_\infty(r)$, is the same as in the r-a model, with parameters $r_0 = 2$, and $k = 3$. Unless otherwise specified, we've assumed for simplicity $w_{ee} = w_{ie} = w_e$ and $w_{ii} = 0$. However, the behaviors are robust to a range of weight values.

Model Implementation

Phase plane and bifurcation analysis of the model in the absence of noise was implemented in XPP, and a similar code was implemented in MATLAB for simulations of the model with noisy input, for the analysis of UP/DOWN state durations. Noise was implemented using Ornstein-Uhlenbeck noise.

$$d\xi = -\theta \xi dt + \sigma\sqrt{2\theta dt} dW$$

with time scale $\theta = 0.05$ and standard deviation $\sigma = 0.25$ unless otherwise specified.

Simulations of equations [1-2] and [3-5] were performed in Matlab using the ode45 solver, with input noise computed using forward Euler method with time step dt=0.1. Accuracy was assessed by comparing results for time steps dt=0.1 and dt=0.05 for a subset of simulations. Statistics for simulations with noise were determined by simulations of duration 60,000 (AU).

A simulated time course was determined to have UP/DOWN states if the distribution of r(t) was bimodal, as determined using a hartigans dip test (Hartigan and Hartigan 1985, implementation at http://www.nicprice.net/diptest/). UP/DOWN state transitions were detected as threshold crossings between high and low rate states. To avoid spurious transition detection due to noise, a "sticky" threshold was used: the threshold for DOWN->UP transitions was taken to be the midpoint between positive crossings of a threshold between the high rate peak of the rate distribution and the inter-peak trough, while the threshold for UP->DOWN transitions was the midpoint between the low rate peak of the rate distribution and the inter-peak trough.

All simulation and analysis code is available at https://github.com/dlevenstein/Levensteinetal2018.

Analytical Solutions for Bifurcation Parameters for the r-a model.

To analyze the parameter space of the model, we used standard procedures from dynamical systems theory (Strogatz). To determine the linear stability and class of a fixed point ($\dot{r} = \dot{a} = 0$) at $[r^*,\ a^*]$, we evaluate the eigenvalues of the Jacobian matrix

$$J = \begin{bmatrix} \frac{\partial \dot{r}}{\partial r} & \frac{\partial \dot{r}}{\partial a} \\ \frac{\partial \dot{a}}{\partial r} & \frac{\partial \dot{a}}{\partial a} \end{bmatrix}_{[r^*\ a^*]} = \begin{bmatrix} -1 + \frac{\partial R_\infty(X)}{\partial r} & \frac{\partial R_\infty(X)}{\partial a} \\ \frac{1}{\tau}\frac{\partial A_\infty(r)}{\partial r} & -\frac{1}{\tau} \end{bmatrix}_{[r^*\ a^*]}$$

where $X = wr - ba + I$ is the total input to the population. Simplifying the partial derivatives as

$$\frac{\partial R_\infty(X)}{\partial r} = \frac{\partial R_\infty(X)}{\partial X}\frac{\partial X}{\partial r} = w\frac{\partial R_\infty(X)}{\partial X}$$
$$\frac{\partial R_\infty(X)}{\partial a} = \frac{\partial R_\infty(X)}{\partial X}\frac{\partial X}{\partial a} = -b\frac{\partial R_\infty(X)}{\partial X}$$

gives the Jacobian for an arbitrary fixed point:

$$J = \begin{bmatrix} -1 + w\frac{\partial R_\infty(X)}{\partial X} & -b\frac{\partial R_\infty(X)}{\partial X} \\ \frac{1}{\tau}\frac{\partial A_\infty(r)}{\partial r} & -\frac{1}{\tau} \end{bmatrix}_{[r^*\ a^*]}$$

where

$$\frac{\partial R_\infty(X)}{\partial X} = \frac{e^{-(X-x0)}}{(e^{-(X-x0)}+1)^2} = r(1-r)$$
$$\frac{\partial A_\infty(r)}{\partial r} = \frac{ke^{-k(r-r0)}}{(e^{-k(r-r0)}+1)^2} = ka(1-a)$$

and thus

$$J = \begin{bmatrix} -1 + wr(1-r) & -br(1-r) \\ \frac{1}{\tau}a(1-a) & -\frac{1}{\tau} \end{bmatrix}_{[r^*\ a^*]}$$

We define $I_{1/2}$ as the level of drive for which there is a fixed point at $r = a = 0.5$, and thus

$$r = R_\infty(wr - ba + I)$$
$$0.5 = R_\infty\big(w(0.5) - b(0.5) + I_{1/2}\big)$$
$$I_{1/2} = R_\infty^{-1}(0.5) + 0.5(b - w)$$
$$I_{1/2} = x_0 - 0.5(w - b)$$

When there is sufficient recurrent excitation for UP/DOWN alternations (i.e. $w > w_0$, see derivation of w_0 below), $I_{1/2}$ gives the level of drive for equi-duration UP and DOWN states, for a given level of recurrence and adaptation strength (Figure S3).

If we define $I^* = I - I_{1/2}$ as the drive relative to $I_{1/2}$, the bifurcation diagrams in I^*-w and I^*-b parameter space (Figure S4A) reveal that $I^* = 0$ (i.e. $I = I_{1/2}$) acts as an axis of symmetry of the effective I/O curve, with Excitable$_{\text{DOWN/UP}}$ regimes surrounding a bistable or oscillatory regime that has equi-duration UP/DOWN states at $I^* = 0$, depending on the values of w, b (Figure 4A, S3A). Furthermore, transitions of the dynamic regime at the center of the I/O curve happen at $I^* = 0$. Figure S4B shows the bifurcations at $I^* = 0$ with changing w and fixed $b = 1$. As w is increased, the fixed point at $r = 0.5$ loses stability with the appearance of oscillations in a Hopf bifurcation at $w = w_0$. With further increased values of w, two stable fixed points appear at high and low rate, marking the transition from oscillations to bistability in a pair of saddle node bifurcations at $w = w_\chi$. Finally, the pair of "inner" unstable fixed points coalesce in a pitchfork bifurcation at $w = w_{PF}$. Thus, the bifurcations at $I^* = 0$ reveal the parameter values at which qualitative changes in the I/O curve occur, between monotonic stable, oscillatory-centered, and bistable-centered I/O curves (Figure 4D).

To solve for the type of fixed point at $I = I_{1/2}$, we use the total input

$$X = wr - ba + I_{1/2}$$

$$X = w(r - 0.5) - b(a - 0.5) + x_0$$

For the fixed point at $[r^*\ a^*] = [0.5\ 0.5]$, $X = x_0$, giving

$$J = \begin{bmatrix} -1 + \frac{w}{4} & -\frac{b}{4} \\ \frac{k}{4\tau} & -\frac{1}{\tau} \end{bmatrix}$$

which we can use to obtain the conditions for w_0 and w_{PF}.

The condition for a Hopf bifurcation at w_0 (Figure 4D, S4) is that J has a pair of purely imaginary eigenvalues $\lambda_\pm = 0 \pm bi$ (Strogatz). $\lambda_\pm$ can be found by

$$\lambda_\pm = \frac{1}{2}\left(\mathrm{Tr} \pm \sqrt{\mathrm{Tr}^2 - 4\mathrm{Det}}\right)$$

where Tr and Det are the trace and determinant of J, respectively. The value for w_0 is found from the null-trace condition:

$$\mathrm{Tr} = 0$$

$$\left(-1 + \frac{w_0}{4}\right) + \frac{-1}{\tau} = 0$$

$$w_0 = 4\left(1 + \frac{1}{\tau}\right)$$

and minimal value for b follows from:

$$0 < \mathrm{Det}$$

$$0 < \left(-1 + \frac{w_0}{4}\right)\left(\frac{-1}{\tau}\right) - \left(\frac{b}{4}\right)\left(\frac{k}{4\tau}\right)$$

$$b > \frac{16}{\tau k}$$

The pitchfork bifurcation at w_{PF} is the transition from 5 (3 unstable) to 3 (1 unstable) fixed points. w_{PF} satisfies the condition that J has a degenerate pair of eigenvalues $\lambda_{\pm} = 0$.

$$0 = \frac{1}{2}\left(\mathrm{Tr} \pm \sqrt{\mathrm{Tr}^2 - 4\mathrm{Det}}\right)$$

$$0 = \mathrm{Det}$$

$$0 = \left(-1 + \frac{w_{PF}}{4}\right)\left(\frac{-1}{\tau}\right) - \left(\frac{b}{4}\right)\left(\frac{k}{4\tau}\right)$$

$$w_{PF} = \frac{bk}{4} + 4$$

The degenerate pair of saddle node bifurcations at w_{χ}, which separates oscillatory-centered and bistable-centered I/O curves, is determined numerically using XPP.

UP/DOWN State Duration Matching

In vivo and simulated UP/DOWN state durations were compared using a non-parametric distribution matching procedure (Figure S5). Similarity was calculated as

$$s = (1 - KS_{UP}) * (1 - KS_{DOWN})$$

where

$$KS_{UP/DOWN} = \sup_x |F_v(x) - F_s(x)|$$

is the Kolmogorov-Smirnov (KS) statistic, in which $\sup_x$ is the supremium function and $F_{s/v}(x)$ are the empirical cumulative distributions of simulated and in vivo durations. In short, KS measures the largest difference between the observed cumulative distributions for simulated and in vivo durations, where $KS_{UP} = 0$ indicates that the in vivo/simulated UP state durations distributions are identical and $KS_{UP} = 1$ indicates that the in vivo/simulated DOWN state durations distributions are non-overlapping. Similarity is thus bounded between 0 and 1, where $s = 1$ indicates that both UP and DOWN state distributions are identical between simulation and the experimental observation, and $s = 0$ indicates that either the observed UP or DOWN state distributions are non-overlapping with the modeled durations.

There is one free parameter in the fitting procedure, which is τ, the population time constant, or equivalently, the time scale factor from non-dimensionalized model time and

seconds. For each simulation, we tested time scale factors from 1ms to 25ms with increments of 0.1ms and used the time scale parameter that gave the highest value for s, thus preserving the shapes of the distributions and the relative values of UP/DOWN state durations.

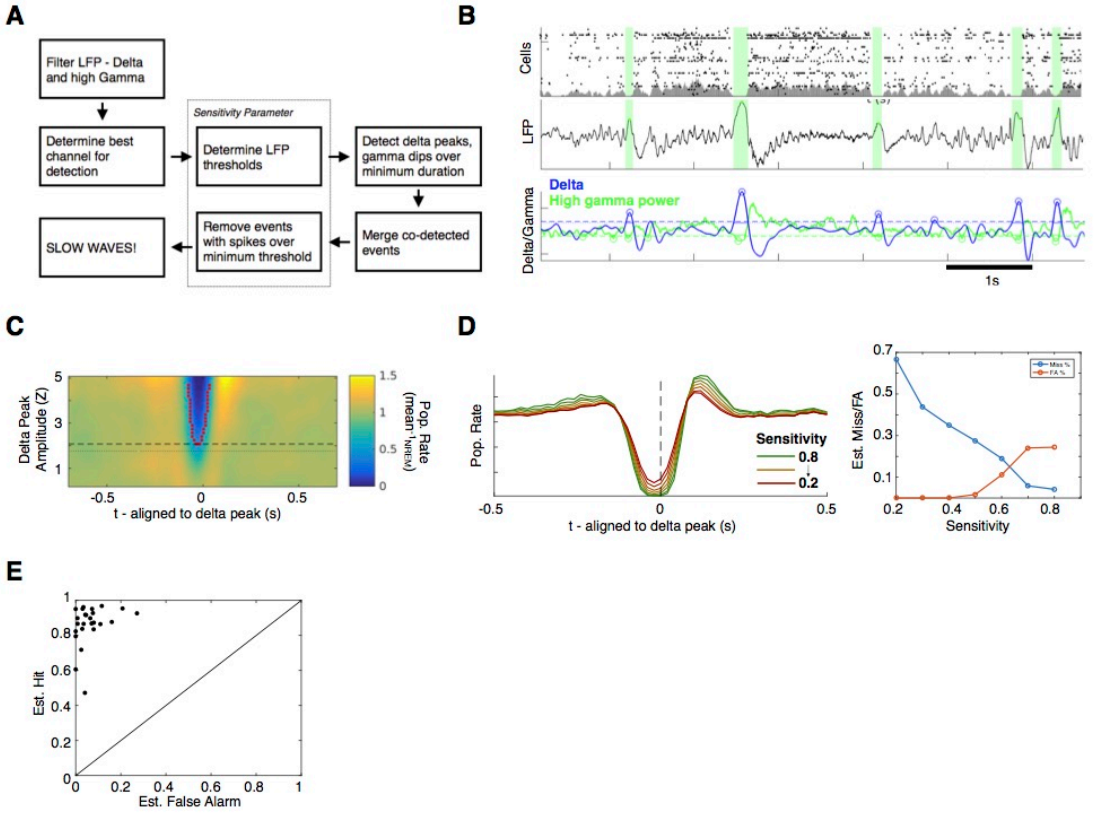

Figure S1 Slow wave detection. Related to Figure 1. **A:** The process for slow wave detection, as described in the Methods, is schematized. **B:** A 7s sample of neocortical data is shown with spike raster (top), LFP (middle), and filtered delta and high gamma power (bottom). Detection thresholds, as determined in panel C, are indicated as dashed (peak) and dotted (window) lines. Slow waves are detected as coincidence of threshold crossings of both signals. **C:** Threshold detection. Slow wave threshold was determined by calculating the population PETH around delta peaks, as a function of peak amplitude. Peak threshold (dashed line) was taken to be the lowest amplitude for which the PETH drops below a (mean-normalized) rate (given as the "sensitivity parameter"). Window threshold (dotted line) was taken to be the average value of the delta signal at which the PETH drops below the sensitivity parameter for any peak value (red dots). **D:** (Left) PETH around detected slow waves in the example recording in B,C for a range of sensitivity parameters. (Right) Estimated miss and false alarm % as a function of sensitivity parameter. Sensitivity = 0.6 was used for this study. **E:** Miss and false alarm % for each recording in the dataset.

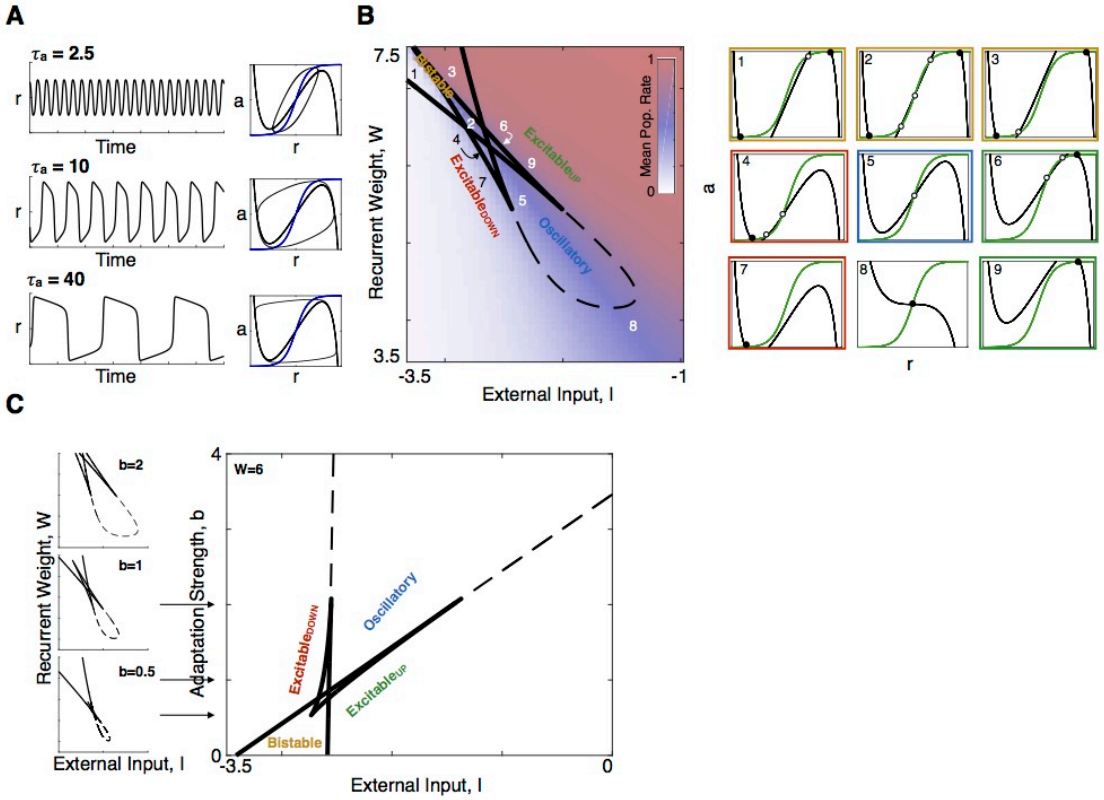

Figure S2 Parameter space of dynamical regimes of the r-a model. Related to Figure 3. **A:** Trajectory of an limit cycle with increasing τ_a. As adaptation gets slower (τ_a gets larger) the limit cycle becomes more "square-shaped" with distinct UP and DOWN states. **B:** The I-w parameter space. Representative phase plane from each domain of parameter space is shown on the right. Regime is determined by the location of stable fixed points. (Yellow: bistable, blue: oscillatory, red: Excitable$_{DOWN}$, green: Excitable$_{UP}$). **C:** The I-b parameter space. Increasing adaptation strength, b, increases the domain of the oscillatory regime. (Left) I-w parameter space for different values of b.

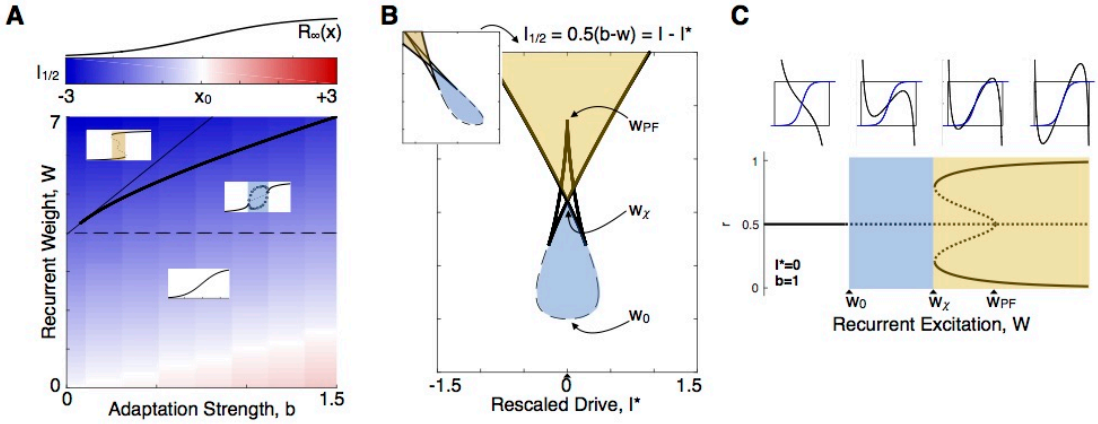

Figure S3 Dynamic regime at the I/O curve center region. Related to Figure 3 **A:** The level of drive drive for half-activation of the effective I/O curve ($I_{1/2}$, see Methods) varies as a function of w and b. Color indicates $I_{1/2}$ relative to x_0 (the midpoint of the unconnected I/O curve $R_\infty(x)$, shown as reference). Lines reflect bifurcations of the effective I/O curve, as shown in Figure 3I, revealing that the operating range for I shifts to lower values (deeper blue) when recurrence is sufficiently high for UP/DOWN alternations. **B:** I*-w parameter space. $I^*=I-I_{1/2}$ centers the parameter space around the I/O curve. **C:** Bifurcations in w, with $I^* = 0$ (i.e. at $I=I_{1/2}$). Oscillations appear in a hopf bifurcation at w_0, two stable (and two unstable) fixed points appear in a pair of saddle-node bifurcations at w_χ, and the unstable fixed points coalesce with the center unstable fixed point in a pitchfork bifurcation at w_{PF}. Each of these bifurcations is shown in the w-b plane in Figure 3I.

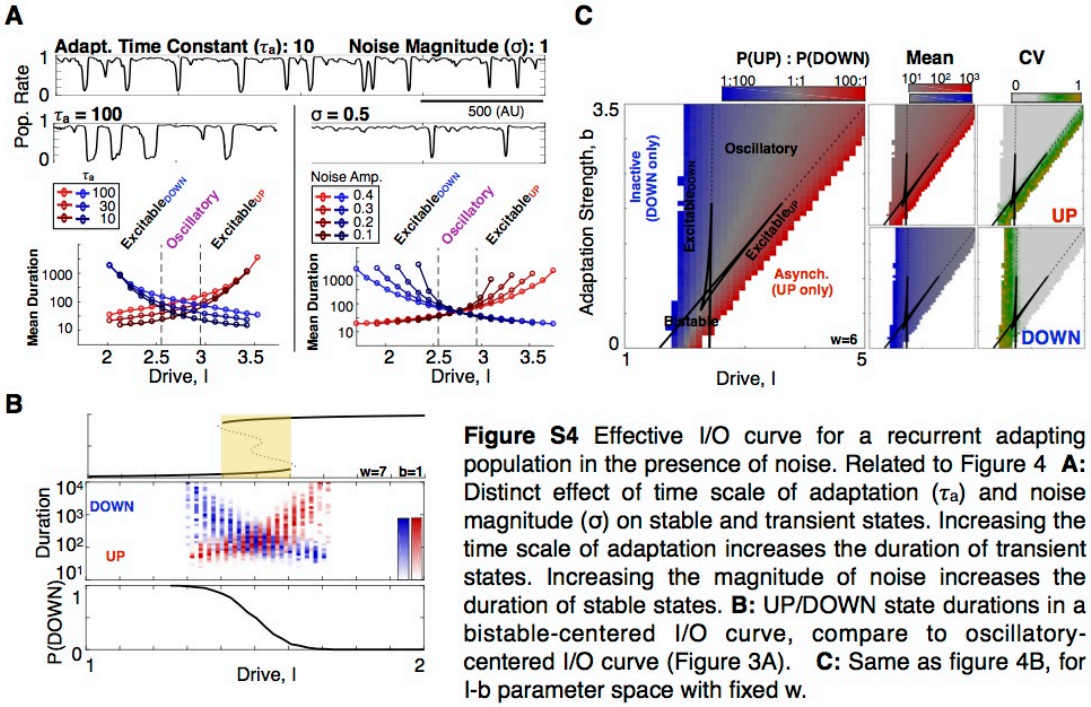

Figure S4 Effective I/O curve for a recurrent adapting population in the presence of noise. Related to Figure 4 **A:** Distinct effect of time scale of adaptation (τ_a) and noise magnitude (σ) on stable and transient states. Increasing the time scale of adaptation increases the duration of transient states. Increasing the magnitude of noise increases the duration of stable states. **B:** UP/DOWN state durations in a bistable-centered I/O curve, compare to oscillatory-centered I/O curve (Figure 3A). **C:** Same as figure 4B, for I-b parameter space with fixed w.

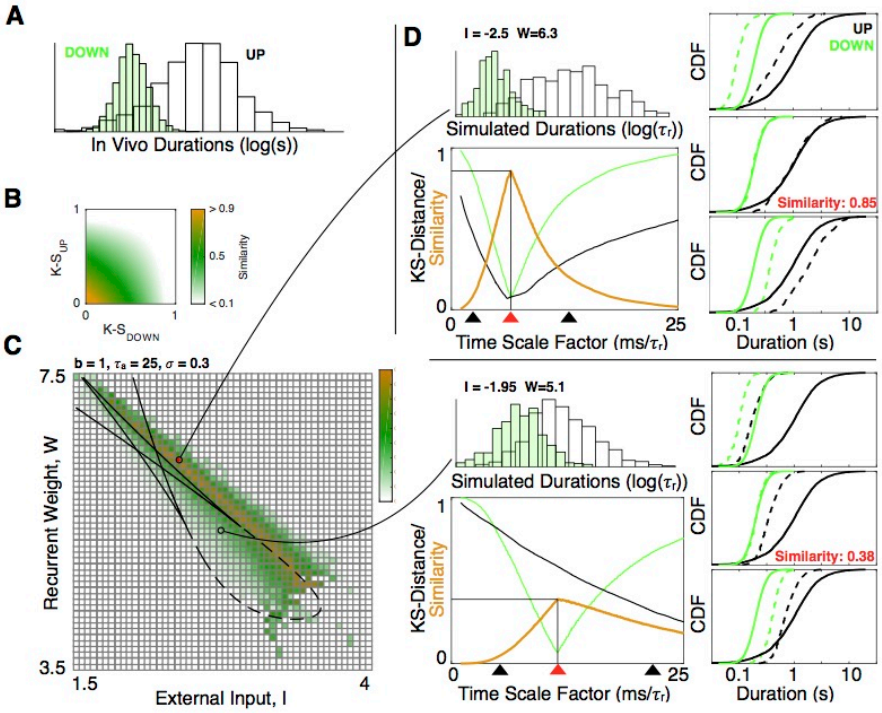

Figure S5 Duration distribution matching method. Related to Figure 5 **A:** UP/DOWN state duration distributions for an example recording. **B:** Similarity as a function of K-S distance between in vivo and simulated UP and DOWN state duration distributions. **C:** Map of similarity between the durations for the example recording in A and the r-a model with noise (Figure 4B). **D:** Calculation of similarity for two points in the I-w plane. For each point, KS distance between simulated and in vivo UP/DOWN state duration distributions is calculated at time scaling factors between 0.5 and 25ms/τ_r. Simulated (dashed) and in vivo (solid) cumulative distributions are shown for three scaling factors (arrows) at each point. Similarity for the point is taken to be the similarity at the best time scale factor, re-dimensionalizing time to give the best match to the shapes and relative values of the UP and DOWN state duration distributions.

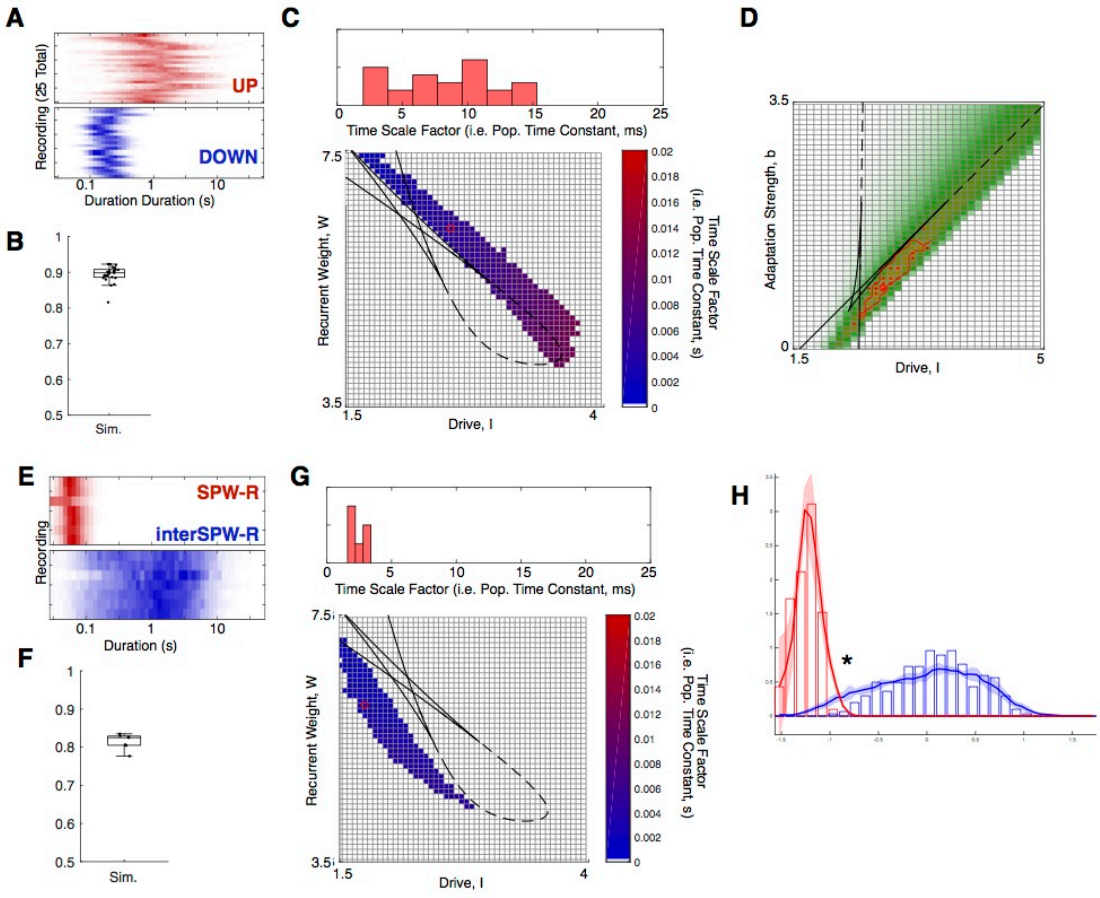

Figure S6 Distribution matching in Neocortex and Hippocampus. Related to Figure 5. **A.** UP/DOWN state durations for all neocortical recordings in the dataset. **B:** Similarity metric for each neocortical recording the dataset. **C:** Histogram of time scale factors for all neocortical recordings in the dataset. (Bottom) mean time scale factor in I-w parameter space. **D:** Matching in the I-B parameter space. Same as figure 5B. **E-G:** Same as A-C, for hippocampus. **H:** Enlarged HPC duration distributions - model does not capture short duration interSWR periods.

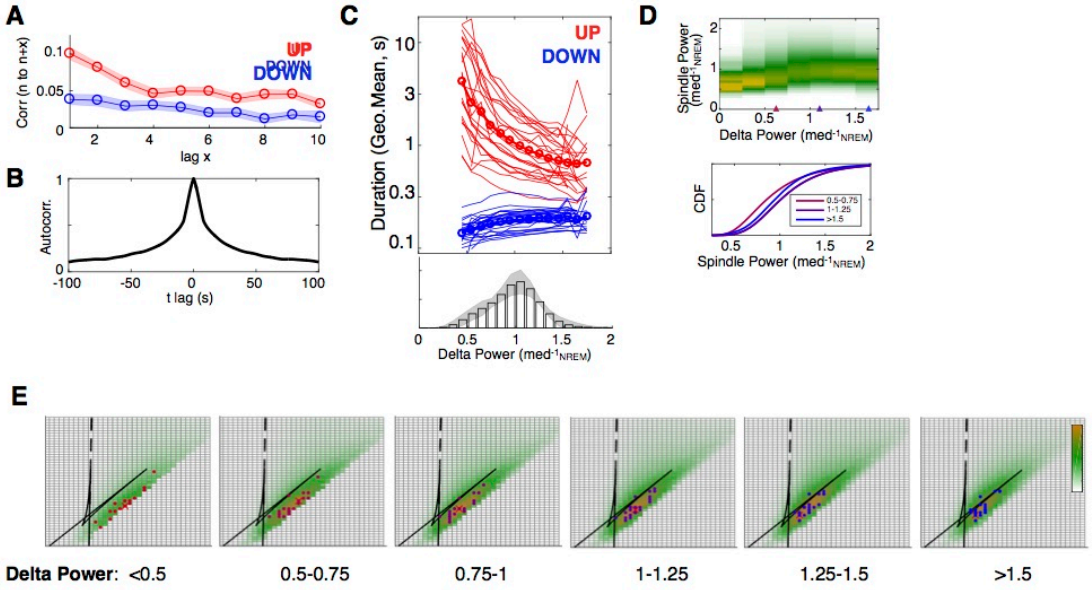

Figure S7 Slow variation in cortical state. Related to Figure 6. **A:** Correlation between durations of adjacent UP/DOWN states. **B:** Autocorrelation of delta power. **C:** Mean UP/DOWN state duration plotted vs delta power. Light lines: geometric mean duration for each recording. Dark lines: geometric mean duration over all recordings. **D:** High spindle power most common for intermediate delta power. Spindle power distribution as a function of delta power over all recordings. (Bottom) cumulative distribution of spindle power for low, intermediate, and high delta power. **E-F:** Same as Figure 6D-E but with matching to the I-b parameter space.

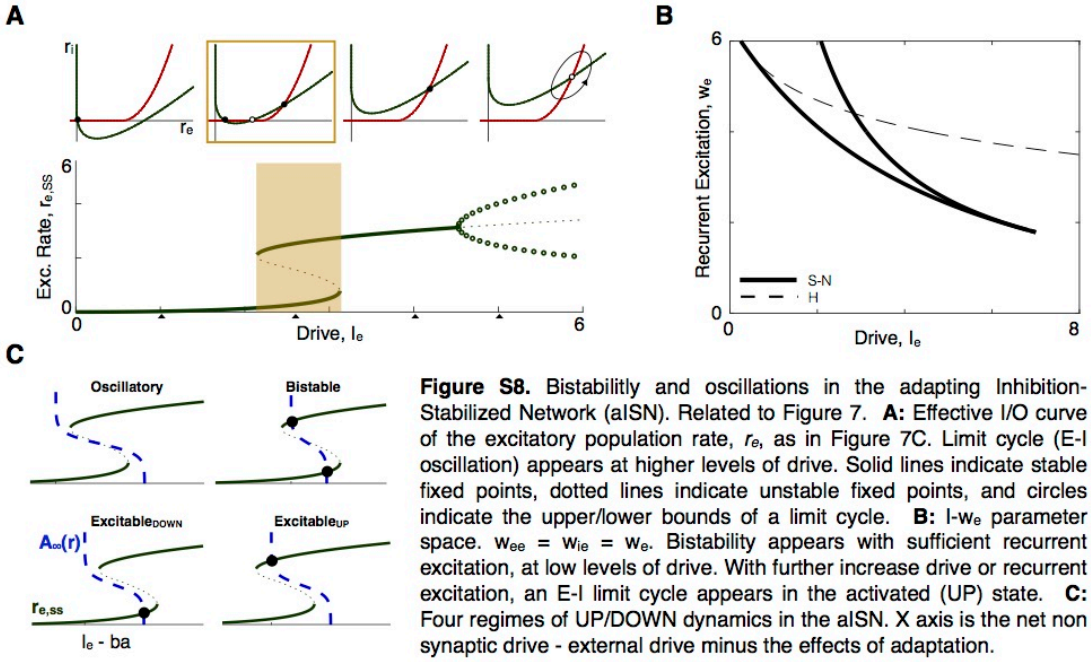

Figure S8. Bistabilitly and oscillations in the adapting Inhibition-Stabilized Network (aISN). Related to Figure 7. **A:** Effective I/O curve of the excitatory population rate, r_e, as in Figure 7C. Limit cycle (E-I oscillation) appears at higher levels of drive. Solid lines indicate stable fixed points, dotted lines indicate unstable fixed points, and circles indicate the upper/lower bounds of a limit cycle. **B:** I-w_e parameter space. w_{ee} = w_{ie} = w_e. Bistability appears with sufficient recurrent excitation, at low levels of drive. With further increase drive or recurrent excitation, an E-I limit cycle appears in the activated (UP) state. **C:** Four regimes of UP/DOWN dynamics in the aISN. X axis is the net non synaptic drive - external drive minus the effects of adaptation.