

# 1 Teoría de la información

La entropía es una medida de cuán «sorprendente» o impredecible es una variable aleatoria. Un patrón regular y predecible es poco interesante, mientras un patrón complejo y de alta impredecibilidad contiene más información. Daremos una medida de entropía para las respuestas neuronales.

Por el momento, vamos a caracterizar las respuestas neuronales temporalmente y en función de su *spike count rate* (SPR), o sea el número de disparos por unidad de tiempo. Definimos  $P(r)$  como la probabilidad de observar un SPR de  $r$ . La entropía de Shannon, que es la medida más común de entropía, expresa cuán «sorprendente» es un valor de  $r$  en términos de su probabilidad, y necesita por lo tanto una función compuesta  $h(P(r))$ . En particular, la entropía de Shannon cuantifica la entropía en términos de la esperanza de  $h(P(r))$ , donde  $h$  se elige de manera tal que satisfaga dos propiedades deseables.

En primer lugar,  $h$  debería ser una función decreciente de  $P(r)$ , de modo que los eventos más probables contribuyan menos a la entropía total que los eventos menos probables. Esto es razonable: si tiramos un dado de seis caras 60 veces y, en general, obtenemos más o menos 10 veces cada cara, no podemos decir demasiado acerca del dado. Pero si obtenemos 30 veces el número 6 y sólo 6 veces cada uno de los otros números, podemos concluir que el dado está sesgado. A esto nos referimos cuando decimos que los eventos menos probables son más informativos.

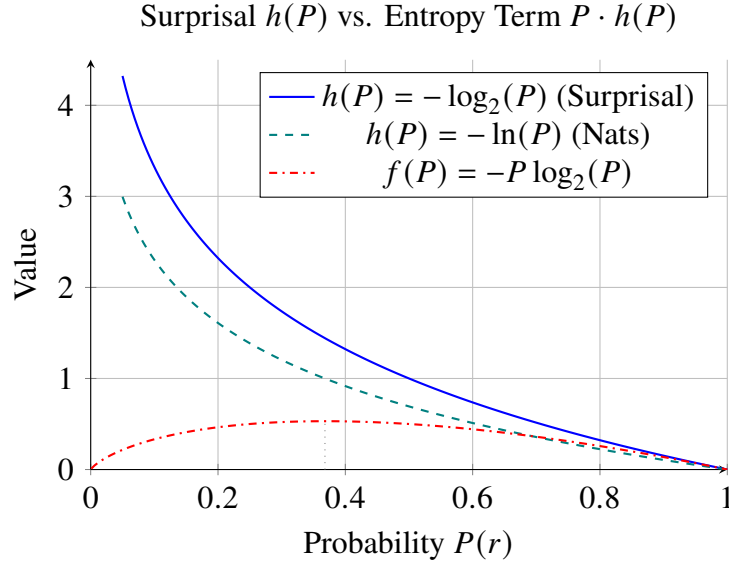
Otra propiedad deseable es que la entropía de dos variables aleatorias independientes debería ser la suma de las entropías individuales. En el caso de respuestas neuronales, si deseamos establecer la entropía de las respuestas de dos neuronas independientes, la entropía total debería ser la suma de las entropías individuales.

El logaritmo es la única función que satisface estas dos propiedades para toda probabilidad  $P(r)$ . Por lo tanto, definimos la entropía de Shannon  $H$  como la esperanza del logaritmo de la probabilidad de las respuestas neuronales:

$$h(P(r)) = -\log_2(P(r)) \quad (1)$$

Usamos el logaritmo base dos por convención, y el signo negativo se añade para satisfacer la primera propiedad, ya que el logaritmo es una función creciente. Habiendo determinado la función que mide la «sorpresa» de una respuesta neuronal, podemos definir la entropía de Shannon  $H$  como la esperanza de esta función:

$$H := \mathbb{E}[h(r)] = -\sum_r P(r) \log_2(P(r)) \quad (2)$$



La entropía de Shannon en última instancia es una medida que depende de la variabilidad, pero no nos dice nada acerca de la fuente de esta variabilidad. Para cuantificar la fuente de la variabilidad usamos la información mutua (MI, por sus siglas en inglés). La idea que la inspira es simple: digamos que un mismo estímulo repetido induce cierta variabilidad  $V_1$  en el SPR, y que una diversidad de estímulos distintos induce otra variabilidad  $V_2$ . Si el SPR es informativo respecto a la naturaleza del estímulo, esperaríamos que  $V_2$  sea mucho mayor que  $V_1$ .

Más formalmente, la MI es la diferencia entre la entropía total, por un lado, y la entropía promedio de respuestas a un mismo estímulo, por otro. Claramente, si MI es cero es porque las respuestas varían tanto con los mismos estímulos como ante el total de estímulos, i.e. no tienen un comportamiento específico e informativo respecto a la naturaleza del estímulo. Definimos entonces:

$$H_s = - \sum_r P(r | s) \log_2 P(r | s) \quad (3)$$

la entropía ante un estímulo particular  $s$ . El promedio de esta entropía sobre el espacio de estímulos nos da la así llamada *noise entropy*:

$$H_{\text{noise}} = \sum_s P(s) H_s = - \sum_{s,r} P(s) P(r | s) \log_2 P(r | s) \quad (4)$$

Esta es la entropía asociada a la variabilidad en la respuesta que *no* depende de cambios en el estímulo, sino que proviene de otras fuentes. Esto debería ser claro:

cada término  $H_s$  es la entropía «dentro» de un estímulo fijo, i.e. la que no depende del estímulo, y  $H_{\text{noise}}$  es la esperanza de este valor. Si la información mutua es la diferencia entre la entropía total y la noise entropy, tenemos:

$$\begin{aligned} I_m &:= H - H_{\text{noise}} \\ &= - \sum_r P(r) \log_2 P(r) + \sum_{s,r} P(s) P(r | s) \log_2 P(r | s) \end{aligned}$$

Usando la ley de probabilidad total, según la cual

$$P(r) = \sum_s P(s) P(r | s)$$

obtenemos

$$I_m = \sum_{s,r} P(s) P(r | s) \log_2 \left( \frac{P(r | s)}{P(r)} \right) \quad (5)$$

La probabilidad conjunta  $P(r, s)$  satisface

$$P(r, s) = P(s) P(r | s) = P(r) P(s | r)$$

Por lo tanto,

$$I_m = \sum_{s,r} P(r, s) \log_2 \left( \frac{P(r, s)}{P(r) P(s)} \right) \quad (6)$$

Esta forma de escribir la información mutua revela que es simétrica respecto a  $s$  y  $r$ : la información mutua que un conjunto de respuestas da respecto de un conjunto de estímulos es la misma que da el conjunto de estímulos respecto a las respuestas.

Veamos algunos casos básicos. Digamos que las respuestas de una neurona no son afectadas por la naturaleza del estímulo, i.e.  $P(r | s) = P(r)$ . Entonces se tiene inmediatamente que  $I_m = 0$ . En el otro extremo, si cada estímulo  $s$  produce una única respuesta  $r_s$  distinta a todas las demás,  $P(r_s) = P(s)$  y  $P(r | s) = 1 \iff r = r_s$ , y  $P(r | s) = 0$  en el caso contrario. Por lo tanto,

$$I_m = \sum_s P(s) \log_2 \left( \frac{1}{P(r_s)} \right) = - \sum_s P(s) \log_2 P(s) = H_s$$

Es decir, la información mutua es exactamente la entropía del estímulo.

## 2 Redes con tasas de disparo

### 2.1 Recordando

Recordemos que

$$\rho(t) = \sum_{i=1}^n \delta(t - t_i)$$

es la *neural response function* (NRF). Si  $t_1, \dots, t_n$  son tiempos de disparo de una neurona,  $\rho(t)$  es 1 si  $t = t_i$  para algún  $i = 1, \dots, n$ , cero de otro modo. El punto es que  $t_1, \dots, t_n$  son valores discretos pero  $\rho(t)$  es una función continua. Usando que

$$\sum_{i=1}^n h(t - t_i) = \int_0^T h(\tau) \rho(t - \tau) d\tau$$

ahora podemos operar continuamente sobre los tiempos de disparo. También se tenía un firing rate  $r(t)$  definido como la probabilidad de obtener un disparo en el tiempo  $t$ .

### 2.2 Presynaptic Current

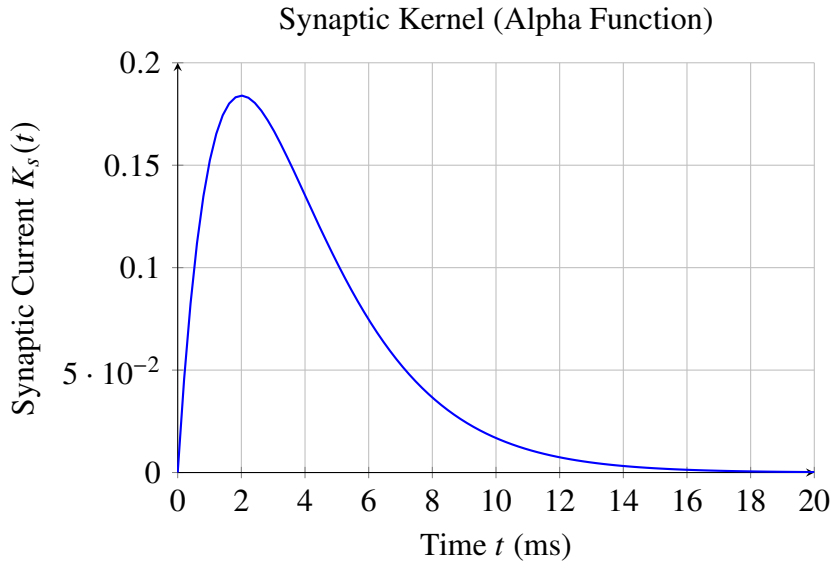
Sea  $\mathbf{u}$  el vector con los *firing rates* de  $N_u$  inputs sinápticos sobre una neurona. Si fijamos el tiempo  $t = 0$  como el momento en que un potencial de acción se dispara en uno de los inputs  $u_k$ , con  $1 \leq k \leq N_u$ , entonces la corriente generada por dicho potencial en el soma de la neurona recipiente puede describirse como

$$I_{\text{generada}} = w_k K_s(t)$$

donde  $K_s(t)$  describe la evolución temporal de la corriente y  $w_b$  su amplitud y sentido. Es decir, se impone que

$$\int_0^{\infty} K_s(t) dt = 1$$

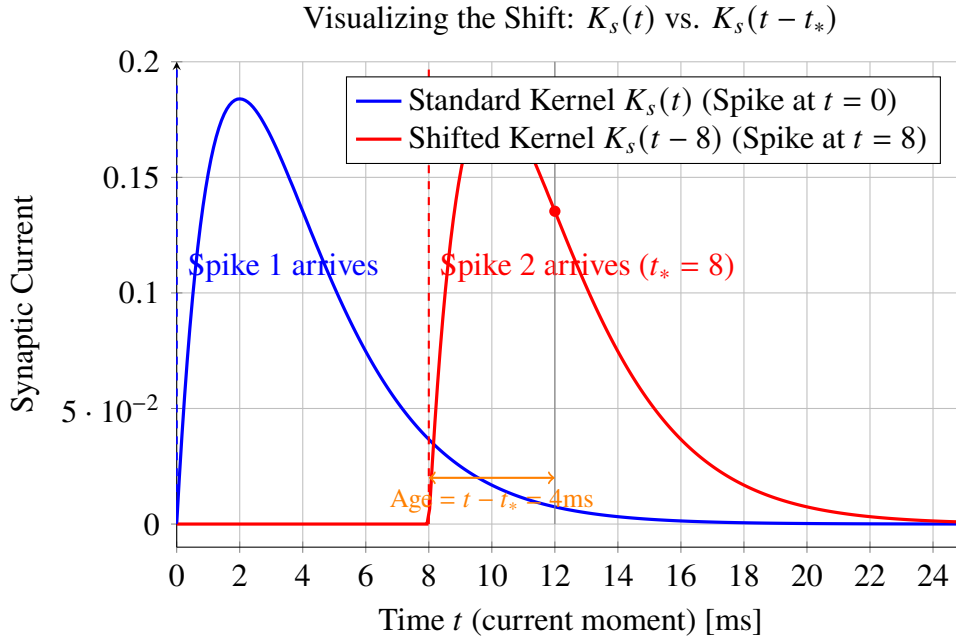
Si  $w_k > 0$  ha habido excitación, si  $w_k < 0$  inhibición. Por ejemplo, un posible modelo para  $K_s$  es la siguiente función: el rápido ascenso muestra la apertura de canales iónicos y demás procesos bien conocidos, con una caída más lenta de la corriente.



Si asumimos que los disparos en una sinapsis específica son independientes los unos de los otros, la corriente total en un tiempo  $t$  resultante de una secuencia de disparos en tiempos  $t_i$  es:

$$w_b \sum_{t_i < t} K_s(t - t_i) = w_b \int_{-\infty}^t K_s(t - \tau) \rho_b(\tau) d\tau \quad (7)$$

La expresión se entiende mejor con un ejemplo. Notemos que el argumento  $t - t_i$  de  $K_s$  es el tiempo transcurrido desde el  $i$ -ésimo disparo y el momento actual. Digamos tenemos una sinapsis lenta que tarda 10ms en alcanzar su pico justo después de un disparo. Esto significa que  $K_s(10)$  es el valor máximo de  $K_s$  y  $K(0)$  es muy pequeño o nulo. Imaginemos ahora que un disparo acontece justo en el tiempo 100ms. En el tiempo posterior 110ms medimos la corriente en el valor  $110 - 100$ , i.e.  $K_s(110 - 100) = K_s(10) = K_{\max}$ . Esto tiene sentido, porque al medir justo diez milisegundos después del disparo, la corriente inducida por el disparo es máxima. Lo mismo sucedería si medimos un tiempo mucho más tardío, obteniendo una corriente menor. Se ve entonces que la ecuación anterior suma las corrientes inducidas por cada disparo anterior al momento actual en el momento actual.



Si la interacción entre las distintas sinapsis es lineal, entonces la corriente total generada por todas las sinapsis en la neurona es simplemente la sumatoria de la ecuación anterior sobre  $b = 1, \dots, N_u$ :

$$I_s = \sum_{b=1}^{N_u} w_b \int_{-\infty}^t K_s(t - \tau) \rho_b(\tau) d\tau \quad (8)$$

Sin embargo, no nos interesa expresar la respuesta inducida en función de los disparos en las neuronas presinápticas, sino en sus tasas de disparo. Por lo tanto, reemplazamos  $\rho_b(\tau)$  por  $u_b$ :

$$I_s = \sum_{b=1}^{N_u} w_b \int_{-\infty}^t K_s(t - \tau) \mu_b(\tau) d\tau \quad (9)$$

Si usamos

$$K_s(t) = \frac{e^{-t/\tau_r}}{\tau_r}$$

entonces

$$\tau_s \frac{dI_s}{dt} = -I_s + \sum_{b=1}^{N_u} w_b u_b = -I_s + \mathbf{w} \cdot \mathbf{u} \quad (10)$$

Es decir,  $I_s$  se describe con una ecuación diferencial.

### 2.3 Post-synaptic current

Si la corriente sináptica  $I_s$  es constante, el firing rate de la neurona post-sináptica se puede expresar como  $v = F(I_s)$ . La función  $F$  se llama activation function y suele tomarse como una función de saturación (e.g. sigmoidea). Más comúnmente, se usa una rectificación de media onda:

$$F(I_s) = [I_s - \gamma]_+ \quad (11)$$

donde  $\gamma \in \mathbb{R}^+$  es el umbral de rectificación. En este contexto,  $\gamma$  modela la propiedad física de las neuronas que hace que disparen si y solo si la corriente supera un umbral específico. Si  $I_s < \gamma$  se cumple  $F(I_s) = 0$ , y si  $I_s > \gamma$  se cumple  $F(I_s) = I_s - \gamma$ , donde el «sobrante» de esta diferencia determina la magnitud de la respuesta. Por convención, la variable  $I_s$  en  $F$  no se toma en unidades de nanoampers sino de Hertz, donde  $\gamma$  también está en Hertz. Quiero decir que  $\gamma$  no es la fuerza que tiene que tener la corriente para que se induzca un potencial post-sináptico, sino la rapidez de la tasa de disparo en la neurona necesaria para inducir un potencial post-sináptico.

Si nos olvidamos del tiempo,  $v = F(I_s)$  ya completa el modelo. La ecuación

$$\tau_s \frac{dI_s}{dt} = -I_s + \sum_{b=1}^{N_u} w_b u_b = -I_s + \mathbf{w} \cdot \mathbf{u}$$

si no hay variación en el tiempo da simplemente

$$I_s = \mathbf{w} \cdot \mathbf{u}$$

Esto genera un output firing rate de

$$v_\infty = F(\mathbf{w} \cdot \mathbf{u}) \quad (12)$$

que nos dice cuánto una neurona responde a una corriente constante.



Otra forma de pensar la tasa de disparo de la neurona post-sináptica es asumir que la misma no responde instantáneamente a cambios en la corriente sináptica. La corriente sináptica puede aumentar o decrecer, pero dicho cambio afecta a la tasa de disparo con cierta fase temporal debida a los cambios estructurales inducidos (apertura de canales iónicos, por ejemplo). Se toma entonces un modelo que es un low-pass filtered version de la tasa de disparo anterior:

$$\tau_r \frac{dv}{dt} = F(I_s(t)) - v(t) \quad (13)$$

En este modelo, el cambio en la tasa de disparo depende de cuánto le falta a la tasa actual  $v(t)$  para alcanzar el estado de equilibrio determinado por  $I_s(t)$ . Si el estado de equilibrio es alcanzado, el cambio es cero. Esta expresión suele escribirse así, pero es lo mismo:

$$\tau_r \frac{dv}{dt} = -v + F(I_s(t))$$

donde  $\tau_r$  determina la rapidez con que el firing rate se aproxima a su valor estable para la constante  $I_s$ , y qué tan precisamente  $v$  sigue las fluctuaciones temporales de  $I_s(t)$ .

Vamos a asumir que  $t_s \ll \tau_r$ , es decir que la corriente  $I_s$  inducida por los disparos presinápticos alcanza su equilibrio muy rápido en comparación con la tasa de disparo de la neurona post-sináptica. Entonces podemos aproximar  $I_s = \mathbf{w} \cdot \mathbf{u}$  en la ecuación anterior. Obtenemos:

$$\tau_r \frac{dv}{dt} = -v + F(\mathbf{w} \cdot \mathbf{u}) \quad (14)$$

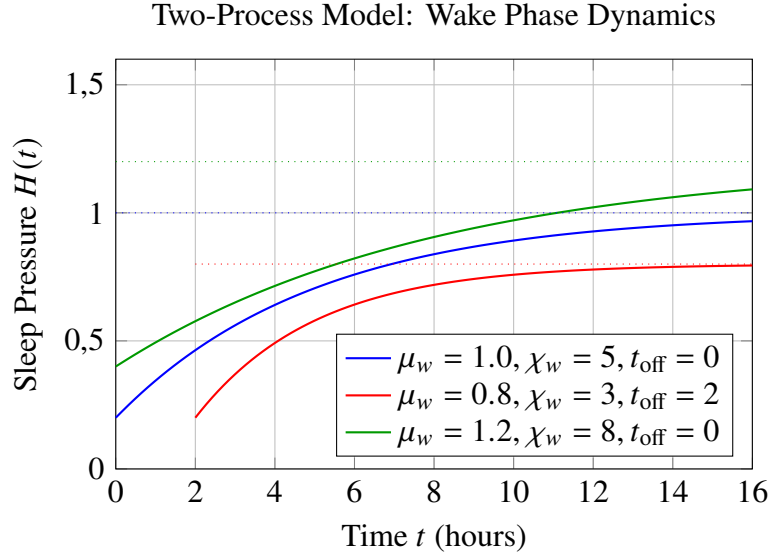
Este es el modelo final.

### 3 The two-process model

The two-process model posits a homeostatic sleep pressure  $H(t)$  which increases during wake and dissipates during sleep. During wakefulness,  $H(t)$  takes the following general form:

$$H(t) = \mu_w + (H(t_{\text{off}}) - \mu_w) \exp\left(-\frac{t - t_{\text{off}}}{\chi_w}\right) \quad (15)$$

where  $H(t_{\text{off}}) < \mu_w$ .  $\mu_w$  acts as the upper asymptote (saturation level). The term  $(H(t_{\text{off}}) - \mu_w)$  is a negative scaling constant. Consequently, while the exponential component decays over time, the overall function  $H(t)$  increases, asymptotically approaching  $\mu_w$  as the 'distance' to the saturation point shrinks.



The model for sleep is the same but with different parameters:

$$H(t) = \mu_s + (H(t_{\text{on}}) - \mu_s) \exp\left(-\frac{t - t_{\text{on}}}{\chi_s}\right) \quad (16)$$

and with  $H(t_{\text{on}}) > \mu_s$ . This makes  $H(t)$  decreasing during sleep. It is easily seen that

$$\frac{dH}{dt} = \frac{\mu}{\chi} - \frac{H}{\chi} \quad (17)$$

It becomes apparent that the change in homeostatic pressure depends on the balance between a constant gain in homeostatic pressure  $\mu/\chi$  and a loss which becomes greater as  $H \rightarrow \infty$ . In short, dissipation is proportional to the instantaneous sleep pressure ( $1/\chi \cdot H$ ), while gain is fixed.

Daan, Beersma and Borbély improved the model by introducing a circadian rhythm  $C(t) = \cos \omega t$  and positive and negative threshold curves:

$$\begin{aligned} H^+(t) &= H_0^+ + \alpha C(t) \\ H^-(t) &= H_0^- + \alpha C(t) \end{aligned}$$

It is now understood that whenever the homeostatic pressure  $H(t)$  coincides with  $H^+(t)$  the upper threshold, there's a transition from wake to sleep. When  $H(t)$  coincides in its decay with  $H^-(t)$ , there's a transition from sleep to wake. Biologically speaking,  $H^+$  represents the resistance to falling asleep, or the maximum pressure the organism can handle before succumbing to sleep.  $H^-$  is interpreted in analogous fashion. No physiological correlates of these thresholds have yet been found, which is a weak point of the model counterbalanced only by the fact that it replicates human sleep patterns quite well. Daan, Beersma and Borbély suggest that these thresholds represent circadian variations in the sensitivity of brain regions involved in the generation of sleep and wakefulness to homeostatic sleep pressure. The upper threshold in particular may be labile to short term environmental, physiological and psychological changes such as light, stimulants (coffee) and motivational aspects [Skeldon and Dijk, 2025].

To analyze the system's behavior purely, researchers often simplify the equations by removing units. This is done by shifting and scaling appropriately. For instance, in the original model, sleep pressure decays down to a lower limit  $\mu_s$ . It is much easier mathematically if things decay to zero. So, we define a new variable  $\tilde{H}$  that represents the *difference* from the baseline:

$$H = \tilde{H} + \mu_s$$

What we do here is equivalent to what is done when speaking of altitude *above sea level*: it's useful to take  $\mu_s$  as reference value and speak only of *sleep pressure above  $\mu_s$* .

By substituting this into our differential equations, the loss term during sleep simplifies. Since the asymptote is now 0, the equation becomes:

$$\frac{d\tilde{H}}{dt} = -\frac{1}{\chi_s} \tilde{H} \quad (\text{during sleep}) \quad (18)$$

The parameter  $\mu_s$  has effectively disappeared. This doesn't mean there is no homeostatic gain during sleep, but rather that now *gain* or *loss* is expressed in the sign of  $\tilde{H}$ .

In analogous fashion, the wake phase is normalized so that the full pressure equals exactly 1. We do this by scaling our variable by the total possible range of the system ( $\mu_w - \mu_s$ ). We update our definition of  $\tilde{H}$  to be a **percentage** of the total range:

$$H = (\mu_w - \mu_s)\tilde{H} + \mu_s$$

Or, rearranged to solve for our new variable  $\tilde{H}$ :

$$\tilde{H} = \frac{H - \mu_s}{\mu_w - \mu_s}$$

Now,  $\tilde{H}$  is a dimensionless number between 0 and 1.

When we plug this scaled variable back into the original differential equations, the constants  $\mu_w$  and  $\mu_s$  cancel out entirely. During wake, the system raises to 1:

$$\frac{d\tilde{H}}{dt} = \frac{1}{\chi_w} - \frac{1}{\chi_w}\tilde{H} \quad (19)$$

During sleep, it falls to zero:

$$\frac{d\tilde{H}}{dt} = -\frac{1}{\chi_s}\tilde{H} \quad (20)$$

The general form during wake becomes:

$$\tilde{H}(t) = 1 + (\tilde{H}(t_{\text{off}}) - 1) \exp\left(-\frac{t - t_{\text{off}}}{\chi_w}\right)$$

The general form during sleep:

$$\tilde{H}(t) = \tilde{H}(t_{\text{on}}) \exp\left(-\frac{t - t_{\text{on}}}{\chi_w}\right) \quad (21)$$

## 4 A physiological model of sleep

Here I present a physiological model developed by Phillips and Robinson in a paper published in 2006. The work is of great interest because previous (and most) sleep models, such as the two-process model, were strictly phenomenological and dealt not with the underlying neurophysiology.

The model is based on the dynamics of the ascending arousal system (AAS), a complex of acetylcholinergic and monoaminergic nuclei in the brainstem. The AAS, which projects diffusely to the cerebrum, controls the arousal state of the brain. Two distinct drives affect the activity of the AAS: the circadian drive and the homeostatic drive, which are previously integrated in the ventrolateral preoptic area (VLPO) of the hypothalamus. Within-system interactions of the AAS—i.e. interactions between the nuclei conforming the AAS—produce the ultradian rhythms characteristic of sleep, like NREM and REM cycles, but are disregarded in the modelling of sleep-wake cycles.

The flow of information across these systems is relatively simple. The homeostatic and circadian drives are both transmitted to the AAS from inhibitory GABAergic projections of the VLPO. The homeostatic signal is caused by the progressive accumulation of adenosine in the extracellular medium of the basal forebrain, which causes the neurons in this region to become inhibited. Since these play an inhibitory role on the VLPO, their own inhibition induces a disinhibition of the VLPO, which itself then signals the AAS more strongly. The circadian signal originates in the suprachiasmatic nucleus and projects to the VLPO via the dorsomedial nucleus of the hypothalamus. It is a fundamentally endogenous oscillator entrained to the day mainly by the influence of light.

In general, it is understood that the monoaminergic nuclei of the AAS promotes wakefulness by forming a mutually inhibitory circuit with the VLPO, which is active during sleep. These two systems are therefore self-reinforcing and conform a flip-flop switch. Contrarily, there is no firm evidence of direct connections from the acetylcholinergic nuclei to the VLPO: it rather seems that the VLPO connects to the acetylcholinergic nuclei via inhibitory interneurons.

Phillips and Robinson's model, hereby termed the PRM, describes three neural populations: the monoaminergic and acetylcholinergic nuclei in the AAS and the VLPO. As described before, the VLPO integrates the homeostatic and circadian drives and inhibits the AAS, an aspect which the model must capture. Since sleep-wake transitions are modulated by the mutual inhibition between the VLPO and the monoaminergic nuclei, the activity of the acetylcholinergic nuclei is modeled at a constant intermediate level. For further justification of this decision, see the original paper.

As most papers, Phillips and Robinson do not provide an intuition behind the mathematical expressions they use. I should wish to provide an explanation not too detailed that it becomes tedious, but sufficient to make the model accessible to, say, an undergraduate student.

First of all, the PRM model does not deal with individual neurons but rather with neural populations. Topologically, the neural populations may be understood as a dynamically evolving weighted graph: each node is a cluster of neurons with some average firing rate and average voltage, with edges representing the sense and magnitude of the interactions between populations. The dynamism of course comes from fluctuations in the voltage, which in turn must affect the firing rate. In general, the relation between the firing rate of a cluster and its voltage is modeled with a sigmoid function, thus capturing the saturation of the firing rate as voltage increases:

$$Q_a(t) = \frac{Q_{\max}}{1 + \exp\left(\frac{V_a(t) - \mu}{\sigma'}\right)} \quad (22)$$

The function  $V_a(t)$  is understood to express voltage relative to the average resting potential of the neurons in the cluster. This means that if  $V_a(t) = 10\text{mV}$  and the mean resting potential is  $-70\text{mV}$ , the actual mean voltage of the neurons in the cluster is  $-60\text{mV}$ . The  $\mu$  parameter defines the mean threshold voltage (also relative to resting). This means that when  $V_a(t)$  crosses  $\mu$ , the firing rate  $Q_a(t)$  is at half its maximum value. Thus,  $\mu$  defines a critical voltage level: below it, most neurons in the cluster are quiescent; above it, most neurons are firing.  $\sigma'$  defines the steepness of the sigmoid around  $\mu$ , and is related to the standard deviation because  $\sigma = \sigma' \pi / \sqrt{3}$ .

The voltage itself is modeled as a leaky integrator of incoming post-synaptic currents into the neurons of the cluster. For the VLPO, which receives and integrates the circadian and homeostatic drives, we have:

$$\tau_v \frac{dV_v}{dt} = -V_v + w_{vm} Q_m + D \quad (23)$$

where  $w_{vm}$  is the weight of the connection from the monoaminergic nuclei to the VLPO,  $Q_m$  the firing rate of the monoaminergic cluster, and  $D$  is the linear combination of the circadian and homeostatic drives:

$$D = w_{vC} C + w_{vH} H$$

As in the two-process model, the homeostatic drive  $H$  is modelled phenomenologically, aiming to capture the metabolic activity that is high during wake and low during sleep. Because arousal correlates highly with monoaminergic activity, the PRM assumes that the homeostatic drive increases proportionally to the firing rate of the monoaminergic nuclei  $Q_m$ . Also as in the two-process model,  $C$  is a simple sinusoidal function with a 24 hour period:

$$C(t) = \cos(\omega t) + \mu_C, \quad \chi \frac{dH}{dt} = -H + \mu_H Q_m \quad (24)$$

Note that this model of  $H$  assumes that when  $Q_m = 0$ , i.e. during sleep,  $H$  is negligible.

The voltage of the monoaminergic nuclei depends on the excitatory input from the acetylcholinergic nuclei and the inhibitory input from the VLPO, from which readily follows the following linear leaky integrator model:

$$\tau_m \frac{dV_m}{dt} = -V_m + w_{mv} Q_v + w_{ma} Q_a \quad (25)$$

The firing rates of the VLPO and the monoaminergic nuclei use the same sigmoid parameters because empirically both populations have similar firing rate dynamics.

As detailed in the original paper, mammalian locus coeruleus neurons have a mean threshold of  $-55\text{mV}$  and a resting potential of  $-65\text{mV}$ , giving a relative threshold of  $\mu = 10\text{mV}$ . Maximal firing rates under prolonged depolarization of single neurons were around  $150\text{sec}^{-1}$ , so we set  $Q_{\max} = 150\text{sec}^{-1}$ .  $\sigma'$  values within  $2 - 6\text{mV}$  for cortical neurons are typical, and Phillips and Robinson argue that because voltage ranges between cortical and brainstem nuclei neurons are similar, a value of  $\sigma' = 3\text{mV}$  is appropriate.

The weights  $w_{jk}$ , representing the strength and sense of the connections from population  $k$  into population  $j$ , have known signs. The VLPO and the monoaminergic populations, as stated earlier, are mutually inhibitory, from which follows  $w_{mv} < 0, w_{vm} < 0$ . The acetylcholinergic population excites the monoaminergic population, so  $w_{ma} > 0$ . The homeostatic drive disinhibits the VLPO, so  $w_{vH} > 0$ , while the circadian drive inhibits it, so  $w_{vC} < 0$ .

In the PRM, the flip-flop switch behavior of the sleep-wake cycle entails an equilibrium with two stable fixed points (wake and sleep) and an unstable fixed point (the threshold between sleep and wake). To find steady states, they solve the system of equations given by setting the voltage derivatives to zero:

$$\begin{aligned}
0 &= -V_v + w_{vm}Q_m + D \\
0 &= -V_m + w_{mv}Q_v + w_{ma}Q_a
\end{aligned}$$

From the second equation, we obtain that the equilibrium voltage satisfies

$$V_m = w_{mv}Q_v + w_{ma}Q_a$$

The firing rate  $Q_m$  at equilibrium is a function of the equilibrium voltage, so that we may write  $Q_m = S(V_m)$ . Thus, from the first equation, it follows

$$0 = -V_v + D + w_{vm}S(w_{mv}Q_v + w_{ma}Q_a)$$

So the equilibrium occurs at the roots of the function

$$\begin{aligned}
f_v(V_v) &= -V_v + D + w_{vm}S(w_{mv}S(V_v) + w_{ma}Q_a) \\
&= \tau_v d \frac{V_v}{dt}
\end{aligned}$$

Phillips and Robinson analyze this function and approximate its roots through the roots of two other, related but simpler functions. These roots are

$$K_{v1} = D + w_{vm}S(w_{ma}Q_{ao}), \quad K_{v2} = D + w_{vm}S(w_{mv}Q_{\max} + w_{ma}Q_{ao})$$

Similarly, they found the equilibrium points for the monoaminergic population. Furthermore, since the firing rate of the monoaminergic and the VLPO populations have the same parameters, they reason that the equilibrium points of both populations should be roughly equal, reducing the dimensionality of the problem. Their mathematical analysis is extensive, so I will simply state that they found the wake equilibrium voltage to fall within  $-15\text{mV}$  and  $-5\text{mV}$ , and the sleep equilibrium voltage to fall within  $0$  and  $1.5\text{mV}$ .