The multiple time scales of sleep dynamics as a challenge for modelling the sleeping brain

BY ECKEHARD OLBRICH1,*, JENS CHRISTIAN CLAUSSEN2 AND PETER ACHERMANN3

1 Max Planck Institute for Mathematics in the Sciences, Inselstraße 22, 04103 Leipzig, Germany
2 Institute for Neuro- and Bioinformatics, University of Lübeck, Ratzeburger Allee 160, 23562 Lübeck, Germany
3 Institute of Pharmacology and Toxicology, University of Zurich, Winterthurerstrasse 190, 8057 Zurich, Switzerland

A particular property of the sleeping brain is that it exhibits dynamics on very different time scales ranging from the typical sleep oscillations such as sleep spindles and slow waves that can be observed in electroencephalogram (EEG) segments of several seconds duration over the transitions between the different sleep stages on a time scale of minutes to the dynamical processes involved in sleep regulation with typical time constants in the range of hours. There is an increasing body of work on mathematical and computational models addressing these different dynamics, however, usually considering only processes on a single time scale. In this paper, we review and present a new analysis of the dynamics of human sleep EEG at the different time scales and relate the findings to recent modelling efforts pointing out both the achievements and remaining challenges.

Keywords: sleep oscillations; sleep stages; sleep regulation

1. Introduction

Sleep is a reversible, periodical state of quiescence and altered consciousness that is more than the absence of being awake—it is an active state that is regulated. The sleeping brain may be considered as a complex dynamical system that is largely independent of external stimulation. Despite the scarcity of external input it shows a rich repertoire of internally generated dynamic patterns and transitions between states. The brain oscillations observed during sleep are complex and span multiple time and spatial scales. Typical changes in the pattern of the electroencephalogram (EEG) serve to discriminate sleep and waking. A first, coarse partitioning is the distinction between wakefulness, rapid eye movement (REM) sleep and non-REM (NREM) sleep.

Sleep appears to be a global phenomenon. Nevertheless, distinct brain regions (cortical and subcortical) are involved in its generation and regulation. For a recent review of important brain structures and neurotransmitter systems relevant for sleep, see Mignot [1].

*Author for correspondence (olbrich@mis.mpg.de).

One contribution of 11 to a Theme Issue ‘The complexity of sleep’.
Figure 1. (a) 20s epoch of EEG data (derivation C3A2) of stage 2 sleep. (b) Colour-coded spectrogram estimated using AR(8) models on overlapping 1s segments (warmer colours denote higher power). (c) Spectrogram of the whole night estimated using AR(16) models on non-overlapping 20s epochs. The black bar denotes the 20s epoch illustrated in (a,b). (d) Hypnogram: MT, movement time; W, waking; REM, REM sleep; 1–4, NREM sleep stages 1–4.

Table 1. Different time scales of sleep dynamics.

<table>
<thead>
<tr>
<th>Time Scale</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>≈ 24h</td>
<td>circadian rhythms and sleep homeostasis</td>
</tr>
<tr>
<td>≈ 90–100min</td>
<td>ultradian process, cyclic alternation between NREM and REM sleep</td>
</tr>
<tr>
<td>seconds to minutes</td>
<td>transitions between sleep stages</td>
</tr>
<tr>
<td>seconds</td>
<td>temporal organization of sleep oscillations, e.g. periodic occurrence of sleep spindles (4s), cyclic alternating patterns (CAPs)</td>
</tr>
<tr>
<td>≥ 1s</td>
<td>slow oscillations, cyclic alteration between ‘up’ (high firing rate) and ‘down’ (no firing) states</td>
</tr>
<tr>
<td>0.05–1s (1–20Hz)</td>
<td>typical sleep oscillations: delta oscillations, sleep spindles, alpha and theta oscillations</td>
</tr>
</tbody>
</table>

Models of the sleeping brain and its dynamics address processes at different levels from the microscopic (cellular) level to the macroscopic (systemic) level and at different time scales from the millisecond or second range up to hours or days.

Phil. Trans. R. Soc. A (2011)
A compact overview on current models is provided by Achermann & Borbély [2]. So far, no model covers all the various aspects. The actual challenge for modelling is to come up with models that integrate the processes on the different time scales.

In this paper, we will focus on the particular problem that the sleeping brain exhibits dynamics on very different time scales ranging from the millisecond (channel opening and closing) to hours or days (sleep–wake behaviour). A rough overview of the time scales considered in this paper is provided in table 1. On a time scale of hours to days the transitions between sleep and waking are determined by the interaction of a circadian process (endogenous clock) and homeostatically regulated sleep pressure $S$ [2]. On an even longer time scale of months and years sleep–wake patterns undergo developmental changes: e.g. from polyphasic sleep–wake patterns in babies to consolidated single sleep episodes in adults to napping behaviour in elderly people.

An additional ultradian process controls the cyclic alternation between NREM and REM sleep with a typical ‘period’ between 90 and 100 min. NREM sleep can be further differentiated into light and deep sleep that are usually scored on 20–30 s epochs [3,4].

A typical feature of the sleeping brain, most pronounced in deep sleep, is the so-called ‘slow oscillation’. The slow oscillation reflects slow membrane potential fluctuations of cortical neurons, which alternate between depolarized up-states (neuronal activity) and hyperpolarized down-states (neuronal silence) on a time scale between 0.5 and 2 s. It gives rise to the slow waves observed in deep sleep and K-complexes in light sleep [5,6]. Other typical sleep oscillations exhibit frequencies between 2 and 20 Hz such as sleep spindles with a frequency of 11–16 Hz. At an intermediate time scale between the occurrence of sleep oscillations and sleep stage transitions some authors have proposed an additional process leading to an alternation between phasic and tonic activity, the so-called cyclic alternating pattern (CAP) [7].

The different time scales are not independent. Sleep stages, for instance, are defined with respect to the occurrence of certain oscillatory patterns [3,4], like slow waves for the definition of NREM sleep stages 3 and 4 or N3. Thus, stage changes are intimately linked with changes in oscillatory activity. Figure 1 illustrates fluctuations of EEG power spectra with a temporal resolution of seconds related to the occurrence of the sleep spindles and changes of EEG power spectra with a resolution of 20 s reflecting transitions between sleep stages and in particular between NREM and REM sleep.

In the following, we will review the dynamics on the different time scales and their interaction in more detail and will present new analyses. Baseline data (sleep from 23.00 to 07.00 h) of a selective REM sleep deprivation study [8] were analysed. Eight healthy young men participated in the study, which consisted of two sessions of nine consecutive nights. In each session, an adaptation night was followed by two baseline nights and six experimental nights. The two sessions were 28 days apart, except for one subject with only 23 days between the sessions. Prior to baseline nights, subjects were engaged in their usual daily activities and were not allowed to nap which was verified by actigraphic recordings. The EEG derivation C3A2 was analysed (sampling rate 128 Hz) and sleep stages were visually scored according to standard criteria [3].
2. Sleep oscillations

Oscillatory activity in the EEG is considered to be an emergent property of the thalamocortical system with the specific patterns and dominant frequencies of these activities depending on the functional state of the brain [9]. During sleep this oscillatory activity can be roughly divided into delta (0.5–4.5 Hz), theta (4.5–8 Hz) and alpha oscillations (8–11 Hz) and sleep spindles (11–16 Hz). Not all of the typical patterns in human sleep EEG are complete oscillations because they may consist only of single half waves. Examples are ‘vertex sharp waves’, K-complexes or the so-called ‘slow oscillation’. Nevertheless, these events contribute essentially to power in the delta frequency range. However, during full blown deep sleep, slow oscillations may become an almost continuous oscillation. The occurrence of these oscillatory patterns is used to distinguish between the different sleep stages. A usual way to detect these oscillatory patterns in the EEG was, and to some extent still is the visual inspection of the data. The most straightforward way to detect events automatically is to apply band pass filters to the EEG according to the aforementioned frequency bands and then use a threshold criterion regarding the amplitude (e.g. through a Hilbert transform) of the activity to detect the corresponding oscillatory activity. The disadvantage of such an approach is that the frequency bands and thus also the corresponding oscillations have to be defined in advance and therefore it is difficult to deal with individual differences or temporal changes of frequencies.

An alternative approach that does not rely on predefined frequency bands was proposed by Olbrich & Achermann [10] and is based on autoregressive (AR) modelling of short (1 s) overlapping EEG segments. The AR($p$) model

$$x(t_n) = \sum_{i=1}^{p} a_i x(t_{n-i}) + \epsilon(t_n)$$

(2.1)

linearly predicts $x(t_n)$ based on previous measurements with $\epsilon(t_n)$ denoting the residual errors. Frequencies $f_k = \phi_k/(2\pi\Delta t)$ and damping coefficients $\gamma_k = 1/\tau_k = -\Delta t^{-1}\ln r_k$, where $\Delta t = t_n - t_{n-1}$ denotes the sampling interval, are estimated from the poles $z_k$ of the AR model which correspond to the eigenmodes of the estimated linear system:

$$z^p - \sum_{k=1}^{p} a_k z^{p-k} = \prod_{k=1}^{p} (z - z_k) \quad z_k = r_k e^{i\phi_k}.$$  

(2.2)

The order $p$ of the model determines the maximum number of poles $z_k$. An AR model of order 8 (AR(8) model) was applied and thus the EEG was modelled as a superposition of maximal four stochastically driven harmonic oscillators with damping and frequency varying in time. Oscillatory events were detected whenever the damping constant at one or more frequencies was below a predefined threshold. A detailed description of the algorithm and its application to human sleep EEG data is provided in Olbrich & Achermann [10]. Figure 2a shows all the detected events in a baseline night as black dots superimposed on the spectrogram. The frequency distribution of the events in figure 2b demonstrates that the events were mainly grouped in three distinct frequency bands, the delta, alpha and spindle bands.
A known mechanism responsible for the generation of oscillations in models of neural systems is the interaction between two populations of excitatory and inhibitory neurons, first studied by Wilson & Cowan [11] and later proposed as a model for the generation of alpha oscillations by Lopes da Silva et al. [12]. For strong enough excitatory–excitatory and excitatory–inhibitory couplings these models exhibit a Hopf bifurcation related to oscillatory activity. Neural field models (NFMs) such as the one of Robinson et al. [13] describe sleep oscillations such as sleep spindles and delta oscillations as resonances near Hopf bifurcations. NFMs are often analysed in the form of ordinary differential equations driven by white noise. If one considers the stationary (fixed point) solution of the deterministic part, the effect of small perturbations can be studied by considering a linear approximation of the model around this fixed point, leading to a structure that can be mapped to a linear time series model. If the Hopf bifurcation is approached the damping of the corresponding resonance decreases and reaches zero at the bifurcation. The closer the state of the system is to the bifurcation, the more pronounced the resulting oscillation in the EEG. Applying the above described algorithm to detect oscillatory events to a simple model system of coupled excitatory and inhibitory neurons revealed that it is possible to detect changes of the eigenmodes of the network resulting from slow parameter changes of the network [14].

(a) Sleep spindles

One prominent sleep oscillation in humans is the sleep spindle. Human sleep spindles are oscillations in the frequency range between 11 and 16 Hz, ideally with a waxing and waning amplitude which resulted in the term ‘spindle’. Their highest incidence is during stage 2 sleep. The occurrence of sleep spindles is one
of the defining features of NREM sleep stage 2 according to Rechtschaffen & Kales [3] and Iber et al. [4]. The spindle density increases in the course of a night [10,15]. This temporal evolution is, in part, owing to a circadian modulation of spindle characteristics [16]. Slow and fast spindles are often distinguished with slow spindles having frequencies between 11 and 12 Hz and fast ones between 12 and 16 Hz, typically resulting in two corresponding peaks in the power spectrum of the EEG [17]. Not all subjects, however, show dual peaks in the power spectrum between 11 and 16 Hz. Moreover, the spindle frequencies vary across the night as a function of sleep depth; they are slower in deep sleep [10] and show large inter-individual differences [17] (see also §4a). Sleep spindles are generated in the thalamus by the interaction between thalamic reticular inhibitory neurons and excitatory cortico-thalamic neurons that project to the cortex [18].

(b) Non-rapid eye movement sleep alpha

In the older literature [19], a third type of sleep spindle was described with an even lower frequency (10 Hz). In the more recent literature, this oscillatory activity is denoted as alpha activity because its frequency is in the range of alpha activity observed in awake subjects resting with closed eyes. In the following, we use the term NREM sleep alpha in order to distinguish the alpha activity during sleep from waking. In a large study comprising 260 subjects only 15 per cent showed clear NREM sleep alpha activity [20]. It is also referred to as alpha–delta pattern [21]. This alpha–delta pattern occurred mostly in the first half of the night with its highest incidence during deep sleep [10]. This activity should not be confused with isolated alpha oscillations that are considered as arousal response [20]. Furthermore, NREM sleep alpha activity shows a homeostatic response, i.e. it is increased after sleep deprivation [22]. In the all-night spectrogram in figure 1 and in the events of figure 2 this activity is evident in the first two NREM sleep episodes with a frequency slightly larger than 10 Hz. This is in contrast with awake alpha activity, seen at the beginning of the recording and approximately 420 min into the recording, which is below 10 Hz.

(c) Slow waves, slow and delta oscillations and K-complexes

Slow waves represent one of the most salient features of the EEG during NREM sleep and reflect sleep intensity. They encompass the frequency range of 0.5–4.5 Hz and were commonly subdivided into slow oscillations (<1 Hz) and delta activity (1–4.5 Hz). Furthermore, K-complexes, a hallmark of stage 2 sleep, contribute to the activity observed in the delta range.

Slow oscillations in the NREM sleep EEG result from slow membrane potential fluctuations of cortical neurons, which alternate between depolarized up-states (neuronal activity) and hyperpolarized down-states (neuronal silence). They were first described in the thalamocortical system of anaesthetized cats [23,24] and later in humans [5,25]. Slow oscillations are generally thought to originate from cortical networks. However, Crunelli & Hughes [26] argued that slow oscillations were emergent properties of cortico-thalamocortical networks, in particular, resulting from a dynamic interplay of cortical, thalamocortical and reticular oscillators [27]. Bazhenov et al. [28] demonstrated that non-homogeneity in extracellular media in a network model may contribute to the generation of slow oscillations.
There is no physiological evidence to draw a sharp border at 1.0 Hz. However, power in the low delta range only partially showed homeostatic behaviour. In particular, power less than 2 Hz does not decline from the first to the second NREM sleep episode [25,29] and was not increased after sleep deprivation [22]. These observations raised the question as to whether slow oscillations are essential for sleep homeostasis.

During sleep stages in which slow waves are predominant, and in the first two NREM sleep episodes, power density spectra showed a peak at 0.7–0.8 Hz [25] indicative of slow oscillations. Negative half-waves predominantly seem to reflect down states [30]. Increased sleep pressure resulted in a redistribution of half-waves between 0.5 and 2 Hz: the number of half-waves per minute was reduced below 0.9 Hz while they were increased above 1.2 Hz. EEG power was increased above 1 Hz only [31]. These findings indicate that slow oscillations contribute to sleep homeostasis and may explain the missing homeostatic response in the power spectrum less than 2 Hz.

Next to slow oscillations, faster delta oscillations (1–4.5 Hz) are observed in the EEG. Their origin is unclear and they may originate in the cortex and in the thalamus [2,32]. In particular, the cortical components are not yet understood. With respect to the functional role of slow oscillations and faster delta activity no clear distinction between these two activities is made in the literature.

Also the contribution of K-complexes is an important issue that needs to be taken into account. They represent a salient feature of stage 2 sleep with a wave form similar to the slow waves that occur during slow wave sleep. They may both be generated by similar physiological mechanisms [5,33] with the K-complexes being the precursors of slow waves [6].

3. The temporal organization of sleep oscillations

Sleep oscillations can either be ongoing oscillatory activity over a relatively long time period or they may appear as isolated short oscillatory events. The first can be applied to slow waves observed during deep sleep, while the latter applies to the occurrence of sleep spindles, K-complexes and to a lesser extent NREM sleep alpha activity in alpha–delta sleep. The typical duration of a sleep spindle is 0.5–1.5 s. With respect to isolated oscillatory events the question arises as to whether they occur according to specific temporal patterns. Evans & Richardson [34] reported a 3–5 s ‘periodicity’ for the occurrence of sleep spindles. Olbrich & Achermann [35] studied whether this finding indicates the involvement of an additional underlying dynamical process or whether it might be explained solely by stochastic processes. Therefore, we compared the inter-event interval distributions of sleep spindles in stage 2 and slow wave sleep with the corresponding distributions of events detected with the same algorithm in data generated by AR models with an eigenfrequency in the spindle frequency range. If the events would occur randomly with a certain rate (Poisson process), we would expect an exponential distribution of inter-event intervals. Owing to the finite duration of the events, we also would expect deviations from this behaviour on short time scales. Two main results were observed (figure 3): (i) the inter-event distributions obtained from simulated stationary processes exhibited a peak between 3 and 5 s similar to the distributions obtained from sleep EEG data and
Figure 3. (a) Inter-event interval distributions of sleep spindles in NREM sleep stage 2 and (b) events in the spindle frequency range detected from data generated by a stationary AR(2) model with $f = 14$ Hz and $r = 0.89$. The dashed curves show fits of exponential functions to the tails of the distributions (see [35] for more details).

Moreover, an interaction between the occurrence of slow oscillations or their presumed precursors, K-complexes and sleep spindles was observed. For example, Mölle et al. [36] found higher spectral power in the spindle frequency range after a K-complex or slow oscillation compared with activity in this band prior to the event, corresponding to the idea that slow oscillations are grouping other sleep oscillations [32].

Several authors [7] made a distinction between phases in which EEG activity alternates between oscillatory patterns (such as K-complexes, bursts of delta waves or alpha activity) and phases without such activity of less than 60 s duration and termed this alternation CAP. It is dissociated from phases longer than 60 s without or with only isolated transient events (non-CAP). Also Achermann & Borbély [25] reported a periodicity of 20–30 s in the prevalence of slow waves. The CAP was considered as some kind of instability of the sleep process. It would be interesting to explore this instability in relation to slow oscillations that were hypothesized to be related to an instability in the cortical firing rate owing to a saddle-node bifurcation in an NFM [37].

Thus, all these different observations indicate that additional dynamical processes on time scales ranging from several seconds up to 1 min may organize the occurrence of sleep oscillations.
4. Sleep stages and sleep stage transitions

In §2, we argued that the EEG activity on short time scales of less than 1s can be parametrized using linear models and that these models also describe most sleep oscillations, with the exception perhaps of slow oscillations. The parameters of these linear models can be interpreted as frequencies and dampings of stochastically driven oscillators. These oscillators can be related to sleep oscillations on the one hand and to resonances of the underlying neural fields on the other hand. The dynamics on the slower time scales therefore corresponds to time dependent changes of these frequencies and dampings. The 4s ‘periodicity’ of the sleep spindle occurrence, for instance, could be related to a corresponding modulation of the damping of the spindle oscillator. Although conceptually appealing there is a problem with this approach: the activity of all oscillators may not be visible concurrently. Distinguishing fast and slow sleep spindles, alpha oscillations, fast and slow delta oscillations and probably also theta oscillations would require one to assume at least six oscillators, which are usually not all visible at the same time. Therefore, a linear model fitted to a short segment of EEG will reveal only one or two of these oscillators, while the properties of the other oscillators cannot be determined from the fitted model. Therefore, the dampings and frequencies of the ‘sleep oscillators’ are practically not useful as features to study their dynamics on slower time scales. An alternative that contains basically the same information as the parameters of the linear models but provides easy to use feature vectors is the power spectrum estimated on short segments.

In order to study correlations between neighbouring frequency bins, we performed principal component analysis (PCA; see also [38] for a related approach).

(a) Principal component analysis

PCA applied to the logarithm of power density spectra during a night is a systematic way to explore the all-night dynamics in the sleep EEG including the sleep stage transitions. In particular, it provides a first estimation of the dimension of the state space needed to describe the slow dynamics by asking how many principal components are necessary to explain, for instance, a certain fraction of the variance of the signal. As feature vectors we used the logarithm of the power density spectra between 0.25 and 64 Hz estimated on non-overlapping 4 or 20s epochs, respectively, with a 0.25 Hz resolution. Figure 4a(i) shows the variances of the first 30 principal components. There were two dominant components and three or four additional components for the 4 or 20s epochs, respectively, until variance changes in a gradual manner. Thus, one could conclude that the space spanned by these spectral feature vectors contains at most six meaningful dimensions. Although the other principal components contain still an important part of the total variance, this variance is almost frequency independent (figure 4a(ii)), and therefore, these components might be considered as representing mainly background noise.

The principal components are characterized by their ‘loadings’ and ‘scores’. The ‘loadings’ show to what extent the single frequency bins contribute to one particular component and the time-dependent ‘scores’ show how much a single component contributes to the power spectrum at the different time points.
Figure 4. (a) PCA applied to power density spectra estimated with a 0.25 Hz resolution on 4 s (blue) and 20 s (red) epochs, respectively of a baseline night (subject VP05). (i) Variances of the first 30 principal components. (ii) Frequency resolved variances of all components (solid) and without the first six components (dashed-dotted). Loadings (iii) and scores (iv) of the first three principal components with the largest variance. Black curves: moving average of the scores over 4 min. (v) Hypnogram based on 20 s epochs (see figure 1 for details). (b) Loadings of the second largest principal component in eight subjects. Each panel illustrates superimposed the loadings determined in four baseline nights of each subject.

We found no qualitative difference between the results of the analysis based on 4 and 20 s epochs, except for larger fluctuations observed in the scores of the first three principal components for the 4 s epochs (figure 4). The loadings of the first principal component are more than 0.05 for frequencies $f > 20$ Hz, except the 50 Hz line frequency. The scores indicate that this first component reflects mainly artefacts and possibly arousals, which may be useful for automated
E. Olbrich et al.

The second principal component showed highest correlations both with sleep stages and with slow-wave activity (SWA). Interestingly, both spindle and delta power entered with the same sign into this loading. This means that this component reflects positively correlated changes in these two frequency bands. By contrast, the third component represents anti-correlated changes between spindle and delta power, as indicated by its loading.

The loadings were relatively stable over different nights of a single person, but differed between persons as illustrated for the second principal component in figure 4b. This result was not unexpected because average power spectra showed trait-like features [39] as do the histograms of the oscillatory events [10].

The black lines in figure 4a(iv) represent moving averages with a sliding window 4 min in duration. The fluctuations around this average represent, at least partially, the processes described in §3. They also illustrate the difficulty in using spectral band power directly as a feature for sleep stage classification. In order to overcome this difficulty, a better understanding of the dynamics on these intermediate time scales is necessary. One needs to keep in mind, however, that sleep stages are not defined solely on the basis of the EEG.

(b) Sleep stages and sleep states

Typical changes in the pattern of the EEG in conjunction with eye movements (EOG) and muscle tone (EMG) serve to discriminate sleep and waking. Mean amplitudes of the EEG, EOG and EMG (high pass filter at approx. 10 Hz) can be determined based on the standard deviation (s.d.) of the signals of consecutive 20 or 30 s epochs (scoring epoch; mean amplitude = $\sqrt{2}$ s.d.). The ratio of EOG/EEG provides a simple measure for the occurrence of eye movements. High values of the EOG/EEG ratio, in combination with low EMG amplitude, are typical for REM sleep episodes. High values of the ratio are also caused by eye movements during waking [40].

Standard rules for scoring the vigilance states in humans were established in 1968 by Rechtschaffen & Kales [3] and recently revised by Iber et al. [4]. These rules define sleep on fixed epoch lengths (20 or 30 s [3]; 30 s [4]). Classically, three basic vigilance states—waking, NREM sleep and REM sleep—are dissociated. NREM sleep is further subdivided into stages 1–4 [3] or N1–N3 [4], with stages 3 and 4 or N3 considered as slow-wave sleep or deep sleep. REM sleep exhibits REMs, which occur phasically under the closed eyelids in conjunction with a loss of muscle tone. Furthermore, autonomic nervous activity shows increased variations during REM sleep that are manifested by fluctuations in heart rate, blood pressure and respiration, as well as by penile erections.

Conventional sleep scoring, however, is inadequate for a quantitative EEG analysis, because the sleep stages are based on rather general and arbitrary criteria and do not reflect gradual changes occurring on a finer time scale. EEG variables are increasingly assessed by computer-aided methods of signal analysis (e.g. spectrogram shown in figure 1, spectral power in specific frequency bands, detection of oscillatory events and PCA).

This raises the question of whether the changes occurring during sleep should be modelled as gradual changes or whether a distinction between ‘macro’ and ‘micro’ states is needed. The classical hypnogram (figure 1) might reflect the macro structure while the occurrence of spindles, arousals, etc., could reflect the

Phil. Trans. R. Soc. A (2011)
micro structure. Obviously, the transition from wakefulness to sleep and vice versa, and the transitions between NREM and REM sleep are state changes. Saper and co-workers [41,42] have proposed a ‘putative flipflop’ for wake–sleep as well as for NREM–REM sleep transitions (see also [43]).

(c) Micro-architecture of sleep

Sleep regulation ensures that sleep states are self-sustained at a time scale of minutes to hours. Nevertheless, deviations in both directions may occur. During wakefulness, consciousness may be lost for a few seconds, as microsleep events occur with increasing sleep pressure resulting in a significant risk for accidents [44]. On the other hand, sleep may be disrupted by short intervals of electrophysiologically identified wakefulness.

Traditionally, sleep interruptions are quantified by the total or average sleep fragmentation, defined as the cumulative duration of interruptions, normalized by the total length of the investigated time interval. Awakenings occur more frequently towards the end of a night [45]. In such an analysis, however, few long interruptions result in the same percentage of sleep fragmentation as frequent short interruptions, but have quite a different impact on sleep quality and recovery. This shortcoming can be overcome by looking at the distribution statistics of the duration of sleep and wake bouts. Such distributions were recently investigated in humans [46] and in mice, rats and cats [47]. These studies reported in all four species two different distributions for sleep and wake bouts. Sleep bouts followed an exponential decay as a function of bout duration, whereas wake bouts by contrast showed a power-law decay over two decades. Lo et al. [46] proposed a phenomenological diffusion model in which the sleep stage range was bounded, and the wake stage range was unbounded with a restoring force corresponding to a logarithmic potential. Apart from the case of a vanishing force (resulting in a power-law distribution with exponent $1/2$) this is the only choice of force that generated a power law in this class of diffusion models.

Lo et al. reported a decay exponent of 2.3 for the power-law decay of wake durations in humans [46,47] and similar values for other mammals. A power law implies the absence of a characteristic time scale. While the sleep bout distribution (in humans) showed an exponential decay of about 22 min (based on 30 s scoring epochs), such a time scale was absent in the distribution of wake bout duration.

In our own dataset scored based on 20 s epochs, we observed similar results (figure 5). In total, 1050 wake intervals were observed. We confirm the exponential decay of the sleep bout durations and observe a decay (based on 20 s scoring intervals) of 14.5 min. According to Lo et al. [47], the decay is dependent on the duration of the scoring interval. This dependence is approximately linear for longer scoring intervals. If we apply the scaling resulting from fig. 3c in Lo et al. [47] we arrive at a correction by a factor of 1.5. Thus, our results are consistent with the 22 min decay reported by Lo et al. [47]. However, one has to keep in mind that these decay time scales depend not only on scoring interval length, species and age, but also on the scoring method applied.

As far as the size of our dataset allows one to conclude, the wake distributions were consistent with a power-law decay with an exponent of $-2.83 \pm 0.15$ (corresponding to $-1.83$ for the cumulative distribution).
Figure 5. Cumulative histograms of wake (dashed curves, squares) and sleep (solid curves, crosses) bout durations during 8h of sleep plotted on (a) a double logarithmic and (b) semilogarithmic scale. Data of four nights of eight healthy subjects were pooled.

plot (figure 5a), the wake bout durations were close to a power law, whereas a power law of the sleep bout durations had to be rejected. Similarly, the semilogarithmic plot (figure 5b) confirmed an exponential decay of sleep bout duration, whereas the wake bout durations did not have an exponential character. How such a pronounced difference between the power-law and exponential distribution could be explained with a physiology-based model still remains an open question.

(d) Age dependence of the wake bout distribution

Lo et al. [46,47] found that the power laws describe wake bout durations of healthy adult mammals in four species well. However, deviations are expected in the case of sleep disorders, under the influence of drugs, and also as a function of age. An interesting study in developing rats [48] showed a clear age dependence in wake and sleep bout distributions: for both distributions a shift towards longer time intervals was found. The distribution of sleep bout duration remained exponential during the investigated age period, whereas the wake bout distribution was exponential in young rats and exhibited a power-law distribution in adult rats [48]. This observation has an important implication for modelling sleep regulation: mechanisms stabilizing wake and sleep states change as a function of age. Such a stabilization seems absent in newborns. Further investigations are needed in particular at advanced age where sleep fragmentation is increased.

5. Sleep regulation

Three distinct processes underlie sleep regulation at the systemic (phenomenological) level: (i) a homeostatic process (process S), whose level is a function of prior sleep and waking; (ii) sleep is also modulated by a circadian process (process C), a clock-like mechanism that is independent of prior sleep and waking; and (iii) an ultradian process occurs within sleep reflected in the cyclic alternation of the two basic sleep states—NREM and REM sleep.
‘Sleep homeostasis’ [2] refers to the sleep–wake dependent aspect of sleep regulation. Homeostatic mechanisms counteract deviations from an average ‘reference level’ of sleep. Thus, sleep propensity or sleep pressure is augmented when sleep is curtailed or absent, and reduced in response to excess sleep.

The pioneering studies of Blake & Gerard [49] revealed that both the arousal threshold and the predominance of slow waves in the EEG are high in the initial part of sleep and then decrease progressively. The occurrence of slow waves in the EEG is quantified by SWA (spectral power in the 0.75–4.5 Hz range), one of the most important functional EEG parameters. Sleep intensity, measured by SWA, is a central part of the concept of sleep homeostasis. The compensation of a sleep deficit occurs mainly by an increase in sleep intensity rather than by the prolongation of sleep duration owing to constraints of the circadian process.

SWA shows a decline in the course of sleep that can be approximated by an exponential decrease across consecutive NREM sleep episodes. The level of SWA at sleep onset increases in a saturating exponential way reflecting the duration of prior wakefulness [50,51]. Thus, the homeostatic process S is modelled with an exponential decline during sleep and a saturating exponential increase during wakefulness.

Slow components in the sleep EEG are thought to underlie the restorative function of sleep. Tononi & Cirelli [52,53] hypothesized that slow oscillations and SWA are reflecting synaptic homeostasis and postulated that synaptic strength is high at the beginning of the night, owing to plastic processes occurring during wakefulness and decreases by means of synaptic downscaling during sleep. Simulations with a large-scale model of the thalamo-cortical system [54] revealed that a decrease in cortical synaptic strength was sufficient to account for the decline in SWA during sleep [55].

The two-process model [51,56] postulates that the homeostatic process S interacts with a circadian process (process C) that is independent of sleep and waking and is generated in the suprachiasmatic nuclei. Process S is assumed to vary between an upper and a lower threshold that are both modulated by a single circadian process. Simulations with this model accounted for diverse phenomena such as recovery from sleep deprivation, circadian phase dependence of sleep duration, sleep during shift work, sleep fragmentation during continuous bed rest and internal desynchronization in the absence of time cues [51].

In a later version of the model [57], the change of process S, and not its level, was assumed to be proportional to the momentary level of SWA. The elaborated model addressed not only the global changes of SWA as represented by process S, but also its time course within NREM sleep episodes. However, REM sleep regulation was not addressed and the occurrence of REM sleep episodes was governed by an external process. In general, a close fit was obtained between the simulated and empirical SWA data and their time course and the occurrence of late SWA peaks during extended sleep could be simulated. The model was recently applied to simulate sleep in narcoleptic patients [58] and Zavada et al. [59] used the model approach to investigate spatial aspects of sleep homeostasis.

Rempe et al. [60] proposed a mathematical model of neuronal substrates underlying sleep–wake regulation. They included NREM–REM sleep alterations based on a previously proposed flip-flop concept [41,42]. Similarly, a model of the sleep–wake dynamics was proposed by Phillips & Robinson [61] and
Robinson et al. [43] with emphasis on the ascending arousal system where circadian and homeostatic influences were integrated. They did not include the ultradian alternation of NREM and REM sleep. Postnova et al. [62] proposed a model in which the homeostatic process is determined by the neuropeptide hypocretin/orexin, a co-transmitter of the lateral hypothalamus. Hypocretin/orexin neurons are silent during sleep and active during wakefulness. During wakefulness synaptic efficacy of hypocretin/orexin declines as a result of the ongoing firing and recovers during sleep. They demonstrated that these homeostatic changes can account for typical alterations of sleep–wake transitions. These approaches have a conceptual resemblance to the two-process model. Their homeostatic component rises (declines) during waking and declines (increases) during sleep. In these models, the homeostatic component, however, has no link to neuronal substrates generating slow waves and is thus not related to SWA.

Further models with various degrees of complexity have been proposed. An overview on current models of sleep mechanisms is provided in Achermann & Borbély [2].

6. Summary and outlook

In this paper, we presented and discussed the dynamics exhibited by the sleeping brain on different time scales as revealed through time series analysis of the human sleep EEG. Furthermore, we highlighted the challenges of modelling sleep on different time scales. For short time scales (e.g. sleep oscillations) and longer ones (e.g. sleep regulation) there exist satisfactory phenomenological, time series and physiology-based models which, however, still can be improved. In particular, the connection between these two time scales needs further elaboration. Furthermore, we raised the question of whether SWA is the correct and only physiological marker for modelling sleep pressure or if other frequency bands also need to be considered. The loadings of the second largest principal component revealed by the PCA indicate that changes in delta, alpha and partially spindle activity might reflect the same regulatory process. Moreover, SWA reflects at least two physiological processes, the slow oscillation (<1Hz) and fast delta oscillations (1–4.5Hz), and the relationship between the slow oscillations and SWA is a complex one, i.e. higher SWA does not necessarily mean more or higher amplitude slow oscillations [31].

Compared with the time scales of sleep oscillations and sleep regulation much less is known about the intermediate time scales with phenomena such as the ‘spindle periodicity’, the CAP or sleep stage changes. At this time scale, further efforts from both time series analysis and physiological modelling are needed. Also the cycling alternation between NREM and REM sleep and the occurrence of short waking episodes (arousals) within a sleep episode deserve further investigations. The ultimate aim should be to develop a model covering the major time scales and being based on a sufficiently detailed biological basis.

We thank Dr Leila Tarokh for comments on the manuscript. The work was supported by the Swiss National Science Foundation (grant 320030-130766 to P.A.), by the Deutsche Forschungsgemeinschaft (DFG) through SFB-654-A8 and Graduate School of Computing in Medicine and Life Science (to J.C.C.).

Phil. Trans. R. Soc. A (2011)
References


*Phil. Trans. R. Soc. A* (2011)


Multiple time scales of sleep dynamics


*Phil. Trans. R. Soc. A* (2011)