Technical advances in the characterization of the complexity of sleep and sleep disorders

Matt T. Bianchi a,1, Robert J. Thomas b,*

a Department of Neurology, Sleep Division, Massachusetts General Hospital, 55 Fruit Street, Wang 720 Neurology, Boston, MA 02114, United States
b Beth Israel Deaconess Medical Center, Department of Medicine, Division of Pulmonary, Critical Care & Sleep Medicine, KB-023, 330 Brookline Avenue, Boston, MA 02215, United States

1. Introduction

Sleep is a complex physiological process that has been linked to the function of every organ system, and dysregulation of sleep has important consequences for health and well-being. For example, obstructive sleep apnea (OSA) has been linked to increased health risks such as hypertension, cardiac disease, and ischemic stroke (Eastwood et al., 2010). However, less obvious disturbances such as too little sleep or even too much sleep have been linked to morbidity and mortality in large epidemiological studies. At the level of daily function, sleep disorders may impact mood, performance, and sleepiness, with concurrent risks of motor vehicle or workplace accidents. Despite these widespread health consequences, and decades of research into the clinical and basic scientific aspects of the sleep–wake system, much uncertainty remains.

The most basic formulation of the sleep–wake system is the two-process model of Achermann and Borbély (2003). In this conceptualization, Process S represents homeostatic sleep pressure that builds in proportion to the amount of time spent awake and drives the system toward sleep. Once sleep ensues, Process S declines, eventually allowing resumption of wakefulness. In parallel, Process C represents the daily circadian rhythm of alertness. The two systems act together to promote sleep consolidation at night, and promote alertness during the day. Although this system began as a phenomenological model, increasing experimental evidence has revealed the biological underpinnings. The suprachiasmatic nucleus governs the circadian rhythm of process C, by interpreting zeitgebers such as light and regulating daily rhythms. Process S may be involved in the buildup of adenosine as a metabolic byproduct of daily waking activity. The two processes can be dissociated experimentally by placing subjects on non-24-hour routines; dissociation also happens culturally, as shift-workers have provided a “natural” experiment demonstrating the effects of forcing wakefulness at times when the system should be sleeping, or forcing sleep during the circadian day time.

The neuro-anatomy of sleep–wake system involves reciprocally connected wake-promoting and sleep-promoting centers (Espana and Scammell, 2004; Saper et al., 2005). The wake-promoting nuclei and their main neurotransmitters include the tuberomammillary nucleus (histamine), the locus ceruleus (norepinephrine), the dorsal raphe (serotonin), and the basal forebrain (acetylcholine). The main sleep-promoting nucleus is the ventrolateral pre-optic (gamma amino butyric acid, GABA). Certain brainstem regions are important for regulating REM sleep, with REM-promoting cholinergic nuclei in the laterodorsal tegmentum and pedunculopontine tegmentum. Neurons in the locus ceruleus, raphe, and tuberomammillary nuclei have correspondingly low activity during REM sleep. Adenosine is also a major sleep-promoting substance, and may act preferentially...
on the basal forebrain regions. The thalamus and in particular the reticular nucleus have been implicated in NREM physiology, including specifically spindle formation. The orexin system of the lateral hypothalamus is thought to provide stabilizing balance between the wake- and sleep-promoting forces. The reciprocal inhibitory connections among these regions have been proposed to function as a so-called “flip-flop” switch, allowing the system to support state transitions between sleep and wake (Fuller et al., 2007).

2. The clinical gold standard: laboratory polysomnography

The gold standard clinical assessment of sleep is the attended laboratory polysomnogram. This includes typically 4 or 6 EEG leads, which forms the basis of much of sleep–wake staging. Electromyography (EMG) of the chin and bilateral electro-oculogram (EOG) complement the EEG and form part of the criteria for identifying REM sleep. EMG of the bilateral anterior tibialis muscles allows the detection of leg movements, most commonly manifesting as ankle flexion mediated by this muscle.

The most recent consensus statement on sleep staging provides for distinctions between wakefulness, REM sleep, and three sub-stages of NREM sleep (Silber et al., 2007). The standard convention for epoch length in clinical studies is 30 s. N1 is the “lightest” stage, consisting of slight slowing of the background rhythm, but absence of classic N2 features such as spindles and K complexes. N3 is characterized by slow waves (~4 Hz) comprising at least 20% of an epoch. REM is characterized commonly by muscle atonia on the chin EMG channel, and the presence of rapid eye movements on the EOG channels. Sleep spindles are ~13 Hz football-shaped oscillations lasting 0.5–1.5 s, and are thought to emerge from thalamocortical loop dynamics (Huguenard and McCormick, 2007). K complexes are large discrete events with initial upward then downward deflection and may have correlates with cortical up–down states (Cash et al., 2009).

Since the pioneering work in the 1930s of Loomis and colleagues to characterize physiological stages, the basic architecture of sleep has remained relatively unchanged. In particular, the Rechtschaffen and Kales characterization of 1968 specified divisions of wake, REM sleep, and 4 stages of NREM sleep (Rechtschaffen and Kales, 1968). NREM stage 1 involved slowing of the EEG relative to wakefulness, but absence of spindles or K-complexes, which characterized NREM stage 2. The proportion of time in which large slow waves were observed determined whether NREM stage 3 (~20–50% of an epoch) or NREM stage 4 (>50% of an epoch) was to be scored. Most recently, in 2007, the American Academy of Sleep Medicine (AASM) guidelines proposed that NREM stage 3 and NREM stage 4 should be combined, and that the NREM stages would thus be re-named N1–N3 (Silber et al., 2007). Although some authors have proposed more fine-grained division of the progression from wakefulness to NREM sleep, these characterizations have not gained acceptance (Santamaria and Chiappa, 1987; Tanaka et al., 1996).

For readers interested in the sleep of animals, the main differences from human scoring is that 10 second (or shorter) duration epochs are more typical, and no distinctions are made among the substages of NREM, all of which is referred to as “slow wave sleep”, in contrast to humans, in which only N3 has this designation.

Other important features include respiration, measured typically by chest and abdomen strain gauge belts, and airflow sensors at the mouth and nose. EKG is also measured routinely, as arrhythmias may emerge during sleep, in particular in association with sleep apnea. All of the above mentioned scoring is currently performed manually by trained technicians, and requires several hours per study. Inter-rater reliability is an ongoing challenge, especially for fragmented sleep patterns. Automated scoring may provide some systematic alternatives, but training a classifier algorithm on a gold-standard that has only modest inter-rater reliability (~80% in some studies) places theoretical limits on performance (Stepnowsky et al., 2004).

Several alternate methods to characterize sleep have been proposed. Measuring motor activation during sleep, typically periodic limb movements (PLMs) can demonstrate complex patterns (Ferri, 2012; Ferri et al., 2012). Fragmented sleep can have excessive motor activation, and excessive motor activation can fragment sleep, but PLMs may be entirely absent in sleep apnea patients and other disorders associated with fragmented sleep. Autonomic approaches such as pulse transit time, peripheral arterial tonometry (PAT) and heart-rate based metrics (variability, transient rate kinetics) provide EEG-free estimates of sleep quality (Bresler et al., 2008; O’Brien and Gozal, 2007; Pepin et al., 2009; Pillar et al., 2003; Tauman et al., 2004). Periods of vigilance are associated with stage N3, but heart rate variability (HRV) metrics are difficult to use when variability is low, such as in heart failure, diabetes, beta-blocker use, or aging. Moreover, the exact temporal borders of periods of vigilance are not well-defined. Cyclic variation in heart rate is a useful marker of sleep apnea but not in those with low HRV. The same limitations apply to Pulse Transit Time — it is difficult to delineate clear periods of “good” vs. “non-good” sleep. PAT characteristics (Bresler et al., 2008) can correlate with conventional slow wave sleep, and may have some utility as a non-EEG marker of sleep quality. Hemodynamic monitoring shows that sleep-related blood pressure reduction (“dipping”) is a marker of health, and adverse effects of non-dipping have a large volume of supportive data (Roulledge et al., 2007; Yano and Kario, 2012). Dipping does not occur abruptly and thus cannot capture short-term changes. Non-dipping can occur in those with good sleep quality by other measures, emphasizing the importance of multi-modal physiological measurements when investigating sleep dynamics.

Respiration characteristics can identify sleep quality. For example, periods of stable breathing dominate sleep-breathing in health and occur for short periods even in those with severe sleep apnea (Jordan et al., 2011; Younes et al., 2012). However, respiratory abnormalities are not present in those with sleep fragmentation from other reasons, such as pain, epilepsy, and auditory noise. Oxygenation and ventilation tracking is useful only when clearly abnormal as sleep quality markers. Endocrine and metabolic markers, such as cortisol and growth hormone levels, cytokines, inflammatory biomarkers, or glucose disposal during sleep, have not been shown to be practical tools to measure sleep quality, in part because of the technical challenges of obtaining such measures outside of specialized research facilities.

Each method of measuring sleep has advantages and limitations, and provides different vantage points into the complexity of sleep physiology. For example, the rich multi-channel data provided by laboratory polysomnogram represents an important window into the mechanisms of sleep and the relationships of the central and autonomic measures of sleep, yet it tells us nothing about longitudinal variations from night to night. Home monitors such as actigraphy, in contrast, exhibit the reverse advantage-limitation pattern, in that longitudinal variation can be observed while the nuances of sleep physiology remain hidden. Each has its place in the armamentarium of the clinician and the researcher.

3. Quantitative analysis of the hypnogram

The most common quantification of sleep stages consists of reporting the sleep efficiency and the percent of sleep time spent in each of the five cardinal sleep–wake stages. Although extensive efforts have been made to evaluate the normative values for these metrics, as well as the impact of medications and diseases, recent work clearly indicates that stage percentage is a strikingly insensitive measure of altered architecture. For example, consider two patients who sleep for 85% of their eight hour time in bed in the sleep laboratory (Fig. 1). Despite having the same sleep efficiency, one had the 15% of wakefulness in a single block at sleep onset, and the other experienced the time awake as multiple one-minute awakenings throughout the night. This striking architecture difference is not evident in
The metric of sleep efficiency. Thus, the most common method of reporting sleep architecture, and indeed the basis of normative amount of each sleep stage, fails to differentiate among arguably quite different phenotypes. One can see in the figure that the bout durations (time spent in any given block of sleep or wake) are quite different in the two examples. Although the average or even median time spent in each sleep–wake stage may provide more information than the stage percentage, these may not capture the complexity of sleep architecture dynamics. Several alternative metrics may be useful to capture this extra information (see below).

The issue of quantifying sleep latency is itself a source of uncertainty (Ogilvie, 2001). For example, in the clinical context, a sleep study may separately indicate sleep onset as the first epoch of any stage of sleep, but other definitions may be seen such as three consecutive epochs of stage N1, or first epoch of stage N2. The “latency to persistent sleep” may be further specified for example as the first epoch of a block of at least 10 min of sleep, to capture the possibility of fragmentation seen at the beginning of observation that might falsely shorten the latency if one of the other definitions is used (imagine a person having just 1 epoch of sleep and then being awake for an hour before entering persistent sleep — different definitions of sleep onset would yield quite different scores in each case). Further complicating matters is the fact that “functional” definitions of sleep onset may yield internally discrepant results, such as the “kill-switch” method (holding a button and awaiting the muscle relaxation that results in button release and thus sleep onset) often showing faster latencies compared to active response methods requiring indication of hearing a tone or other stimulus, and both may differ from EEG-based scoring. The interested reader is directed to the extensive review of this interesting literature by Ogilvie (2001).

The importance of accurate characterization of sleep–wake stages cannot be overstated; one wonders how many studies have reported “no effect” or even escaped publication due to unrevealing analysis of stage percentage, when in fact sleep architecture effects were indeed present. Sleep apnea represents a tangible example of the insensitivity of percentage metrics, and the risk of false negative analysis of sleep architecture when percentages are the focus. Several groups have shown that the stage percentages are quite similar between groups with versus without sleep apnea, even in the case of severe disease (Bianchi et al., 2010; Norman et al., 2006; Swihart et al., 2008). We have also shown that distribution analysis captured the differences associated with performance of yoga, which were not evident in percentage analysis (Kudesia and Bianchi, 2012). In each of these studies, the clearly evident fragmentation was easily demonstrated by analysis of the stage distributions.

Importantly, other metrics that are fairly simple to obtain are in fact sensitive to fragmentation (of any cause theoretically). These include analysis of sleep–wake stage transition frequency and the distribution of sleep–wake bout durations. In the hypothetical examples of Fig. 1, both the transition frequency and the bout distributions are affected even when the stage percentages are not. Accurate quantitative analysis is important because potentially meaningful effects of drugs or diseases can be missed or trivialized if one limits analysis to stage percentages. This is especially important given the known publication bias against so-called negative findings.

While the percentage of time spent in a given sleep–wake stage may be normally distributed, it is clear that sleep–wake bout lengths are not normally distributed. Unfortunately, most research studies do not show bout durations as distributions; they instead present sleep as percentage of time spent in sub-stages or as the mean duration spent in any sub-stage. One clue in reviewing the literature is that large standard deviations may suggest non-normal distributions, especially for small bout lengths (as may be seen for N1 or wake stages in particular). Understanding the distribution of time spent in sleep–wake stages has practical implications for sleep architecture analysis, in that standard parametric tools (means and standard deviations) are inappropriate. The assumptions behind these tests render them insensitive for detecting potential differences in non-normal data sets. Even the non-parametric rank tests may not capture the dynamics, necessitating other strategies.

Although there is debate as to whether sleep–wake bout duration distributions are exponential, multi-exponential, or power-law distributed, it is clear that the dynamics governing stage transitions are complex and worthy of further interrogation for the potential to yield mechanistic and/or biomarker insights of clinical relevance (Blumberg et al., 2005; Chu-Shore et al., 2010; Lo et al., 2004). To investigate the distribution of bout lengths, there are several approaches. One is to plot a histogram showing the frequency of observations in each “bin” of time. This technique has the advantage of being an intuitive way to visualize data, but has several important disadvantages, including the need to specify the bin length and intervals — both of which can affect the interpretation of the data. For example, making the bins too small results in every bin having either zero or one observation (too sparse to understand the structure of the data), while making the bin size too large blurs the underlying structure of the data. Linear binning results in over-sampling of brief events and under-sampling of long events, while non-linear binning is not straightforward and can substantively bias subsequent fitting and analysis. Linear binned histogram data can be useful in that the pattern can be fitted with non-linear functions such as the sum of exponentials, as we have previously reported (Bianchi et al., 2010), or as survival curve analysis as performed by others (Norman et al., 2006). In this way, the shape of the histogram can be compared among groups of patients, by comparing the parameters of exponential fitting. Other methods that are not sensitive to bin size choices or nested best-fit function questions include cumulative probability distributions and survival curves, which can be analyzed in a number of ways.

In our studies, the binned histogram distribution of each sleep and wake stage was separately fitted with the sum of exponential functions. This is accomplished by first fitting with a single exponential

Fig. 1. Stage percentage is an insensitive measure of sleep architecture. From top: the cartoon hypnogram shows an example of continuous sleep following a block of wakefulness before sleep onset. Time is on the X-axis (unlabeled), and the stages of Wake, REM, and NREM sleep are indicated on the vertical axis. The sleep efficiency (percentage of time in bed consisting of sleep) is ~85%. The lower hypnogram shows a highly fragmented night of sleep, characterized by many brief awakenings scattered throughout sleep. Note the sleep efficiency of this hypnogram is also ~85%, yet the dynamics are entirely distinct.

Please cite this article as: Bianchi MT, Thomas RJ. Technical advances in the characterization of the complexity of sleep and sleep disorders. Prog Neuro-Psychopharmacol Biol Psychiatry (2012), http://dx.doi.org/10.1016/j.pnpbp.2012.09.017
decaying function, which is akin to a survival curve but has a particular parametric property, defined by the time constant of decay, which refers to the time required for the decay function to decay to 1/e, or ~36%, of its initial value. Thus, the smaller the decay constant, the faster the function decays, and the shorter the bout durations on average. The next step is to determine whether the addition of a second exponential function improves the quality of the fit to the histogram, and so forth adding extra exponentials one at a time and performing a statistical test (such as an F-test) to determine the necessity of the extra exponential at each step. Each exponential function implies increasingly complex control over the dynamics of time spent in a given state.

Because of uncertainty regarding the proper type of equation to fit the distribution of time spent in sleep and wake stages, we undertook a simulation study to specifically address the question of whether a power law or a multi-exponential function can fit the types of distributions often observed in sleep architecture analysis (Chu-Shore et al., 2010). Power law distributions are also known as “scale free”, an important feature of some complex systems, meaning that there is no typical scale of observation. Practically, this is often taken to mean that the histogram distribution of events, when plotted on log-log axes, appears to be a straight line. However, despite this simplicity, it is well understood that such a pattern (linear on a log-log plot) is not sufficient to declare a function a power-law, nor can one use the typical least squares method of goodness of fit to assess fits of such functions.

To demonstrate these issues and investigate the capacity to mistake a power-law for a multi-exponential and vice versa, we turned to simulations. We therefore generated known power-law distributed sample data sets and fitted their distributions with multi-exponential functions, and likewise we generated known multi-exponential distributions and fitted them with a power-law function (Chu-Shore et al., 2010). Through these Monte-Carlo simulations, we showed that multi-exponential distributions can mimic a power law function, and that this potential for mimicry was most prominent near the time constants and relative proportions of the multi-exponential patterns observed in actual human sleep stage data. We concluded that the patterns of bout durations followed a more probabilistic dynamic rather than a scale-free system dynamic suggested by the power law. Importantly, we also noted the challenge of unambiguously determining for an empiric data set (such as observed in humans) whether the power law or multi-exponential approach is “best”, from a statistical standpoint, even using sophisticated model choice statistics. Without a strong a priori hypothesis, distinguishing between these non-nested functions was not feasible.

The finding that the bouts of sleep and wake stages could be fit with the sum of exponential functions is interesting from a biological standpoint in terms of the dynamics of how long a person remains in a given stage before making a transition out of it. Drawing from the literature of enzyme kinetics and ion channel kinetics, one can infer the existence of “functional states” based on the number of exponential functions. For example, consider by analogy a simple ion channel existing in only two states, open or closed, with a single time constant determining the transitions from open to closed, and another from closed to open. Watching such a channel bounce between these two states, one can collect a distribution of bout durations of open times and of closed times. Each distribution histogram will be described by a single exponential decay function, specified by the rate constant of the transition. In this simple example, the transition connectors between the states are uniquely defined: with only two states, they each much connect uniquely to the other. By extension, if one observes a new ion channel, and the open state distributions are best fit by the sum of two exponentials, and the closed state distributions are best fit by the sum of three exponentials, then one can infer a state diagram that has five total states (two open states and three closed states). The challenge of how they are connected to one another in this case is not trivial (see below). Nevertheless, the important point is that even though only two “phenotype” states are observed, that is, open versus closed, one can infer more than one “generator” state based on the distribution of durations of time spent in each observable state. By analogy to sleep and wake, when we observe multiple exponential functions in the histogram shapes, one can relate these functions to “generator” states in a transition model of sleep architecture.

The use of transition probability models of sleep–wake dynamics has been proposed for three decades, but they are not commonly implemented due in part to computational requirements and the need for more than one night of sleep (which is the typical clinical sleep study duration) for appropriate analysis. For example, we showed that the multi-exponential dynamics of human sleep–wake bout durations was observed best when 300 subject nights were combined (Bianchi et al., 2010). Performing transition analyses with limited data sets is not advised, especially since the variance typical in sleep–wake bouts makes the resolution of multi-scale dynamics difficult to resolve. This is especially true of the longer duration bouts (i.e., the more stable bouts), which occur less frequently, but may represent important aspects of sleep physiology. Another point of caution involves the time scale over which transitions are smoothed or averaged. For example, in standard AASM scoring, the minimum bout duration is 30 s. If one scored sleep over even shorter bout lengths, brief awakenings such as what we currently call arousals will interrupt sleep, causing a relative increase in short duration sleep bout durations and a concomitant decrease in long-duration bouts. Conversely, smoothing the sleep architecture by considering only bouts of minimum duration of 60 s or longer will have the opposite effect — namely, reducing the frequency of brief bouts while concomitantly increasing the duration of longer bouts.

One of the intriguing aspects of human sleep is that the governing processes can yield markedly different sleep–wake stage patterns from night to night. There are some common themes, such as the average of 90 min between REM blocks, and the propensity for more slow wave (N3) sleep to occur in the first half of the night. However, these trends are not strictly observed on individual nights, but rather are observed when sleep data is averaged across nights. The night-to-night variability may have many contributors, including daytime behaviors such as napping, caffeine, nicotine, alcohol, prior night sleep, stress, exercise, and meal times. However, it is also likely that sleep architecture has intrinsic variance, or stochastic behavior, contributing to nightly variability. Stochastic variability is consistent with our finding that the distribution of sleep sub-stages, and of wake durations, follows multi-exponential dynamics (Bianchi et al., 2010). This observation allowed us to build first-order Markov state transition models based on a combination of the exponential function time constants and their relative contributions to the distributions (Bianchi et al., 2011), as described above. We found that wake bouts required three exponentials, NREM sleep bouts required also three exponentials, while REM sleep bouts required two exponentials, in a group of adults without sleep apnea who underwent a single night of home sleep testing. From this, we inferred that the three “phenotype” states (wake, NREM, REM) were generated or governed by a total of eight kinetic or generator states from a modeling perspective. To determine how these eight states actually connect in terms of transition probability, we analyzed the preceding and subsequent bout durations. In other words, with three phenotype states, we can form six pairs of transitions (wake to REM, REM to NREM, NREM to wake, etc.). From these pairs, we can analyze the bout duration histogram of each state contingent upon its adjacent neighbor (such as, the distribution of all REM bouts that ended by a transition to NREM sleep). This provides a set of six exponential fits, the time constants of which constrain which generator states allow transitions to which other generators. From this, we constructed an eight-state model, constrained by empiric observation, which generated realistic sleep hypnograms in the sense that they capture the fragmentation, and the stochastic aspects of “real” sleep architecture. In other words, night-to-night variability was embedded in the model structure, through the probabilistic transition rate constants. Of note, to enable REM cycling, we did have to introduce time variance to the REM-entry transitions.

Please cite this article as: Bianchi MT, Thomas RJ. Technical advances in the characterization of the complexity of sleep and sleep disorders, Prog Neuro-Psychopharmacol Biol Psychiatry (2012), http://dx.doi.org/10.1016/j.pnpbp.2012.09.017
In addition to providing a quantitative and generative model of sleep in the absence of OSA, we were also able to repeat this process for a group of patients with severe OSA. We found that both the time constants and the connectivity of the model generator states were different between patients with versus without OSA. We do note the limitation that the study was performed at the group level, and thus application of Markov transition models to individuals must await the capacity for long-term home monitoring to generate sufficient data to estimate these parameters on a single-person basis for phenotyping. Such analysis theoretically could provide more intricate sleep phenotype fingermarks than the current standard. In the meantime, considering aspects like transition rates and bout distributions (either parametric or non-parametric) should be employed to avoid the false negative issues described above associated with standard stage percentage measures.

4. Quantitative analysis of EEG signals

The slow wave oscillations during NREM sleep have garnered widespread attention due to evidence for involvement in synaptic plasticity as well as a non-invasive biomarker of sleep homeostasis (Crunelli and Hughes, 2010; Massimini et al., 2009; Tononi, 2009). For example, bidirectional changes in slow wave sleep dynamics were observed contralateral to the arm involved in either a motor learning task or an immobilization task during the subsequent night of sleep (Huber et al., 2004, 2006). Complementing this data, which suggests slow wave dynamics react to daytime experience, another study showed that memory could be enhanced by direct current stimulation at low frequency during sleep (Marshall et al., 2006).

The basis for the slow wave oscillation is thought to derive from cortical neuronal dynamics, although subcortical modulatory contributions may occur. Invasive recordings have shown that cortical neuron transitions between depolarized and hyperpolarized membrane potentials, known as “up” and “down” states, is associated with distinct firing patterns (Cash et al., 2009; Destexhe et al., 2007; Holcman and Tsydyn, 2006). Although increased synchrony between adjacent cortical neurons is commonly felt to be required for generating the large amplitude EEG voltage changes of slow wave sleep, there may be a simultaneous decrease in larger scale cortical connectivity. Human subjects undergoing TMS during sleep demonstrated evidence of decreased cortical connectivity nearly immediately upon entry into NREM sleep (Massimini et al., 2005). TMS has also been used to trigger classic-appearing waveforms during NREM sleep, such as slow waves and sleep spindles (Massimini et al., 2007). This intriguing data raises the possibility that non-invasive, non-pharmacological methods may one day become feasible strategies for improving sleep in patient populations.

NREM slow wave activity may subserve important aspects of sleep-dependent neuronal plasticity. Tononi has proposed a synaptic homeostasis hypothesis for slow wave sleep, in which synaptic strengthening occurring during waking experience undergoes nightly pruning to prepare for the next day experiences (Tononi, 2009; Tononi and Cirelli, 2006).

Another technique to extract network dynamics from non-invasive measures involves network analysis drawn from the field of graph theory. EEG or MEG sensors are considered nodes and the correlation between pairs of sensors defines the connections in the graph interpretation of the functional dynamics. This technique gained widespread notoriety after the landmark 1998 publication of Watts and Strogatz (1998) that defined small-world network motif as a special intermediate structure between regular or lattice graphs (that only have local connections) and randomly connected graphs. Small-world network architecture may be enhanced by direct current stimulation at low frequency during sleep (Marshall et al., 2006).

Small world parameters can be extracted from routine 20-lead EEG data. Ferri et al. studied the functional connectivity patterns in healthy adults across different EEG frequency bands during wake, NREM and REM sleep (Ferri et al., 2007, 2008). The clustering coefficient increased during NREM sleep in frequencies below 15 Hz, and this finding was more prominent during periods of cyclic alternating pattern (Gilmartin and Thomas, 2004; Terzano and Parrino, 2000). Although the implications of such findings for sleep disorders remains speculative, the persistence of structured dynamics observed in wakefulness is consistent with studies mentioned above in regard to default, executive, and attention networks.

Scammell and colleagues demonstrated an alternative method for quantifying the sleep EEG, known as state-space analysis (Diniz Behn et al., 2010). By plotting brief epochs of time (2 s) as ratios of power in certain empirically determined frequency bands yielded scatter plots that clustered according to manual staging of wake, REM, and NREM sleep. This method may offer deeper insights into sleep architecture by expressing the phenomena in a more dynamic manner than the traditional sleep staging. The technique can potentially be applied to human EEG data in healthy and disease populations, raising the possibility of quantitative phenotyping.

5. Metabolic imaging

Early studies of brain metabolism using PET and SPECT provided fundamental assessments of brain metabolism (and, by inference, neuronal activity) across wakefulness, NREM sleep and REM sleep (Maquet, 2000). These studies demonstrated that NREM sleep was associated with decreased metabolism and CBF relative to wakefulness. Within NREM sleep, deeper stages were associated with decreased overall metabolic activity and more brain regions demonstrating lower metabolism.

Functional MRI has been combined with EEG and techniques to render the scanner less noisy to facilitate sleep within the scanner (Lovblad et al., 1999). Decreased frontal activity and increased occipital activity were observed during REM sleep relative to NREM sleep. Subsequent studies investigated the response to auditory stimulation during wake versus sleep using combined EEG and fMRI methods (Czisch et al., 2002; Portas et al., 2000). Entry into NREM sleep was associated with decreased auditory fMRI responses. This change was however dependent on the nature of the sound: the response to simple beep sounds was more attenuated than the high-valence sound of the subject’s own name. This finding is consistent with the idea that the brain continues to perform some level of interpretative processing of environmental stimuli even during sleep.

Although these studies have been informative as to the potential substrates of sleep stage physiology, they remain indirect measures of neuronal activity, and the relationship of metabolic changes to neuronal firing is not always straightforward. For example, a neuron could change its pattern of firing, while maintaining the same average rate of firing, and thus presumably encode new information or perform new processing without altering oxygen or glucose consumption. The capacity to extend metabolic imaging to patient populations may be complicated by comorbid diseases (such as diabetes or atherosclerosis) or the use of medications (such as adrenergic modulators) that affect neurovascular coupling within the intracranial circulation. Finally, experimental data from animal studies in which invasive electrode recordings are paired with metabolic imaging have not always shown expected relationships (Schridde et al., 2008), and the BOLD fMRI signal is affected by multiple factors (Birn et al., 2006; Kleinfeld et al., 1998; Shmueli et al., 2007; Wise et al., 2004). Nevertheless, this area of inquiry is under active investigation and holds great promise in the effort to use...
non-invasive imaging methods to infer the neuronal activity associated with healthy and pathological brain activity (Adamantidis et al., 2007; Boyden et al., 2005; Carter et al., 2009, 2010).

The excitement surrounding the discovery of the default mode network has extended to the study of sleep physiology. During quiet wakefulness, in the absence of any task, brain metabolism exhibits correlated activity patterns in a circuit consisting of the posterior parietal and precuneus region, the medial prefrontal cortex, and the lateral parietal regions (Buckner et al., 2008). This pattern dissolves upon task engagement, leading to the idea that the brain has an idling activity pattern, from which it must disengage to attend to tasks. In one small study of 5 healthy adults, simultaneous high density EEG and fMRI measurements revealed persistence of the default network into NREM sleep, which mainly consisted of stage N2 (Larson-Prior et al., 2009). This finding contrasted the hypothesis that generalized network disconnection would be associated with the transition from conscious awareness to sleep, which has been suggested by transcranial magnetic stimulation data (Massimini et al., 2005). However, the findings were consistent with prior work that had already shown default network activity in human NREM sleep (Horovitz et al., 2008), as well as during pharmacological anesthesia (Greicius et al., 2008; Vincent et al., 2007).

Perhaps most surprising was the persistence of wake-related networks, such as the dorsal attention network and the executive control network, in those who have severe sleep fragmentation from pain. The effectiveness of the motor system is lost in periodic limb movement disorder. By integrating more than one subsystem, the confidence of state detection is improved. If two sub-systems (e.g., autonomic + EEG, hemodynamic + respiratory) are ineffective, it is highly likely that other subsystems are also similarly engaged.

Computing coupled and coherent oscillations may provide a view of sleep unconstrained by the limitations of any single system. Pairs of signals include, for example, heart rate variability and respiration, EEG and blood pressure, respiration and blood pressure, and heart rate and blood pressure — all are plausible coupled systems amenable to computational analysis. One method developed uses the electrocardiogram (ECG), from which is extracted autonomic and respiratory influences, both of which are intensely modulated by state (sleep and wake). The resulting “sleep spectrogram” is a map of coupled oscillations during sleep, which yields unique insights into physiological and pathological sleep. A second method uses the EEG and respiratory signals to compute metrics of electrocortical-respiratory coupling (Chervin et al., 2004). It is likely that other methods will be developed.

Effective sleep consists of a certain pattern of linked oscillations among multiple aspects of physiology, such that during periods of effective sleep, generally all components of the sleep system are in a desirable mode, such as stable sleep and breathing, blood pressure dipping, normal oxygenation and ventilation, and absence of EEG arousals. Effectiveness (or ineffectiveness) can however be concordant or discordant. Sleep apnea is an example of concordant ineffectiveness — where all sleep sub-systems can be linked in pathological low-frequency (every 25–40 s) coupled oscillations: apnea/hypopneas, blood pressure surges, heart rate acceleration and deceleration, oxygen desaturations and periodic limb motor activation. However, effectiveness can be considered across other relevant sleep subsystem. As examples, effectiveness of oxygenation and ventilation can be impaired despite normal REM sleep in hypoventilation syndromes. Or, oxygenation can be normal in those who have severe sleep fragmentation from pain. The effectiveness of the motor system is lost in periodic limb movement disorder.

6. Autonomic metrics

Cardiac metrics continue to provide an important autonomic window into sleep physiology. For example, increased heart rate has been associated with insomnia, consistent with the hyper-arousal theory as a common theme in this spectrum of disorders (Bonnet and Arand, 2010). Heart rate variability may reflect the balance between parasympathetic and sympathetic tone. HRV analysis of insomnia patients during sleep suggests excess sympathetic tone (Bonnet and Arand, 1998).

The use of cardiac metrics is promising for the assessment of patients with insomnia for several reasons, including the ability to measure autonomic dynamics over time (given night to night variability) and in the comfort of home (given the sensitivity of sleep to the laboratory environment).

7. Sleep effectiveness

A new concept proposed is that of “sleep effectiveness”. The term effectiveness is used to distinguish it from “efficient” (as the term sleep efficiency has a specific meaning in sleep science) and “fragmented” (as there are numerous opinions, time-scales and definitions of sleep fragmentation). The terms “restorative” and “non-restorative” are already used in the literature but definitions are unclear, as there have been no convincing correlations of these terms with standard sleep staging metrics. Effective sleep is conceptualized as a sleep state that allows the normal functions of sleep (for brain and body; a desirable sleep state to be in), and ineffective sleep as a state that does not. An individual can be “efficient and ineffective”, meaning that their standard metrics are “normal” but the sleep is not restorative, as well as “inefficient and effective”, if the sleep is of high quality, even if the proportion of sleep in relation to time in bed is reduced. The concept is not restricted therefore to traditional sleep stages; rather it is a bimodal characterization of sleep. A good example of how effectiveness can diverge from efficiency is in delayed sleep phase syndrome — efficiency may be low due to increased sleep latency, but once sleeping, the quality of sleep (and thus its effectiveness) can be normal (Fig. 3).
high, low, and very low frequency coupling (Fig. 1). A subset of low frequency coupling is elevated-low frequency coupling (e-LFC), the detection of which requires that the minimum low frequency power be greater than 0.05 normalized units and that the low to high frequency ratio be >30 to define periods of probable apnea/hypopnea or fragmented sleep (Thomas et al., 2007).

High frequency coupling is reduced and low frequency coupling increased in states of fragmented sleep, including depression (Yang et al., 2010), fibromyalgia when conventional polysomnographic stages and arousals did not clearly differentiate controls and disease (Thomas et al., 2010), sleep apnea (Thomas et al., 2005), and heart failure (Yeh et al., 2008). Sleep spectrogram biomarkers are heritable

Fig. 2. ECG-derived sleep effectiveness. From top: 1) Absolute delta power 0–4 Hz (arbitrary units). Note the higher absolute Delta power in the first half of the night compared to the second half. 2) 0–4 Hz delta power normalized to total EEG power. Note that the first vs. second half of night differences is reduced. 3) The logarithm of the ratio of high frequency to low frequency cardiopulmonary coupling. Note the correspondence between delta power fluctuations and the CPC coupling ratios. 4) The cardiopulmonary coupling sleep spectrogram. Note the high correlation ($r=0.84$ in this example) between HFC power and normalized delta power, across the entire night.

Fig. 3. Blood pressure and sleep effectiveness. From top: conventional stages (REM sleep in blue), intra-arterial systolic and diastolic blood pressure, and the sleep spectrogram. Note that dipping occurs in association with high frequency coupling (arrow) even though conventional sleep is stage N2. Thus, the ECG-spectrogram provides improved detection of a sleep state with desirable hemodynamic features.
Movements are stored in 30 weeks outside of the constraints of the laboratory (Ancoli-Israel et al., 2012). High frequency coupling could be considered an integrated biomarker of sleep effectiveness, capturing biological activity in several sleep sub-systems (Thomas and Mietus, 2011).

ECG-derived effectiveness may offer specific advantages when assessing the effects of medications on sleep. For example, benzodiazepines increase high frequency coupling (unpublished data) and the EEG-correlate non-cyclic alternating pattern; this class of drugs also reduces delta power and thus diminishes or eliminates amplitude-defined slow wave sleep. In this instance, conventional sleep staging can suggest detrimental effects, while the ECG-spectrogram may show improved sleep quality. In aged populations, where absolute delta power during sleep is reduced, the ECG-spectrogram can provide unique information about the quality and effectiveness of sleep.

EEG power in the 0.5 to 4 Hz is used as a marker of homeostatic sleep drive. While high frequency coupling occurs in both stages N2 and N3, there is a relationship with relative delta power. The ebb and flow of sleep homeostatic drive across the night are thus reflected in simultaneous changes in cardiopulmonary coupling. Blood pressure decreases during the sleep period; this phenomenon of “dipping” is considered a sign of autonomic health. Dipping occurs only during sustained periods of high frequency coupling, providing an ECG-spectrogram biomarker of a desirable cardiovascular regulatory state (Fig. 2).

7.2. Respiratory cycle related EEG changes, or RCREC

One additional method to describe the effectiveness of sleep is to measure respiratory coupled EEG changes. This technique was developed by Chervin and colleagues at the University of Michigan (Chervin et al., 2004), and seems to provide information beyond that obtained by conventional polysomnographic assessments. In the RCREC method, the respiratory cycle is broken down into early and late inspiration and expiration, and the variance of the typical sleep EEG spectral energy bands across these segments is computed and averaged, the target being non-apneic breaths. Thus, the technique assesses the impact on electrocortical activity of partially obstructed airflow, and the associated physiological consequences (i.e., increased respiratory effort, afferent stimulation from the upper airway, intrathoracic pressure and volume changes, blood gas changes). Correlations of specific signal bands with clinically important outcomes such as objective sleepiness have been shown (Chervin et al., 2005), as well as the effects (reduction in RCREC) of positive pressure therapy (Chervin et al., 2012). RCREC could be generated from corollary rostral-oriented discharges from the respiratory neuronal groups, cranial transmission of peripheral respiratory sensation, and outflow from the solitary nucleus that can integrate chemoreceptor, baroreceptor, and other visceral afferents. The effects of medications, and change over time with treatments, sleep apnea or sleep targeted, are important aspects of this biomarker that need assessment. RCREC could provide a new window into understanding the remodeling of cortical dynamics over time.

8. Movement tracking

Building on the simple observation that wakefulness often involves relatively greater movement than sleep, wearable accelerometers have been utilized to track activity patterns over days to weeks outside of the constraints of the laboratory (Ancoli-Israel et al., 2003). This technique, known as actigraphy, typically involves wearing a wrist-watch device containing multi-axis accelerometers. The watches may also contain light sensors as an additional signal. Movements are stored in 30–60 second bins, and analyzed according to standard algorithms to allow estimation of wake time (based on a certain magnitude and duration of movements) versus sleep time (based on the relative paucity of movement).

Longitudinal data such as this can be used qualitatively to assess circadian rhythm disorders, such as delayed circadian phase, advanced circadian phase, or irregular sleep–wake syndrome. Standard quantification of rest periods can be performed as well, including sleep efficiency, total sleep time, and number of movement periods interrupting the rest period. Increased occurrence of movements during the rest period can be an indication of sleep fragmentation. The circadian “amplitude” can be calculated as the ratio of average activity during active periods divided by the average activity during rest periods. Blunted circadian amplitude can be an indicator that rest periods are fictil, or that wake periods are more sedentary, such as the pattern that may be observed in the elderly. The specification of “sleep” in this methodology is based entirely on the absence of movement, and does not distinguish well quiet wakefulness from sleep, nor does it allow for distinction of different stages within sleep.

More sophisticated analyses are gaining increasing interest as potential biomarkers of sleep physiology. Hu and Shea have utilized a technique of detrended fluctuation analysis (DFA) to examine patterns of activity across time scales in people with or without dementia (Hu et al., 2009). The relation of variance in movement to the time scale at which it is measured (minutes to hours) resembles a power-law distribution in healthy older adults. However, age-matched patients with dementia had an altered relationship, deviating from the power-law pattern, suggesting that the degenerative process involved disruption of rest-activity patterns. Although it remains to be seen whether this sort of finding will prove clinically useful, it does raise the interesting possibility that actigraphy analysis may add to the phenotypic characterization of dementing disorders. The fact that the abnormal patterns were seen in individual subjects with dementia is highly encouraging, as findings limited to group analyses are substantially limited in terms of potential applicability to individual patients in the clinical arena.

Another promising technique was recently proposed by Lim et al., in which they used transition analysis to quantify sleep in elderly subjects using a large database of wrist week-long actigraphy measures from participants in the Rush Memory and Aging Project (Lim et al., 2011). Through dichotomous staging into rest versus activity, they showed that transition dynamics reflected increased fragmentation with age, even after correcting for demographic and medical factors. They also showed that day-to-day variability was high, emphasizing the importance of longitudinal measures in the objective phenotyping of sleep–wake abnormalities. Although it remains to be seen whether this sophisticated transition analysis of traditional actigraphy proves useful from a prospective or biomarker standpoint, the ease of acquisition and automated nature are important features when it comes to generalized application to large datasets.

9. Future directions

The advances in the measurement of sleep together with analytical developments hold promise for improved patient care at multiple levels. For example, it is possible that improvements in characterization of sleep phenotypes will improve the ability to objectively diagnose patients with diseases that are currently based entirely on subjective report, such as insomnia. Improvements in this regard then raise testable hypotheses regarding disease pathophysiology (such as links between hyper-arousal and insomnia, or mapping imaging findings to phenotypes) as well as treatment (such as predicting response to medical or behavioral interventions for insomnia). The application of the novel methods outlined in this review have wide implications, offering the opportunity to learn more from the enormous amount of data that is routinely collected in clinical PSG, as well as focused prospective studies using repeated measures of sleep in individuals with a variety of sleep disorders. These converging themes herald an exciting future for the
clinical practice as well as research advancements in the field of sleep medicine.

References


