Published in final edited form as:

Rep Prog Phys. 2013 September ; 76(9): 096601. doi:10.1088/0034-4885/76/9/096601.

The physics of functional magnetic resonance imaging (fMRI)

Richard B Buxton

Department of Radiology, University of California, San Diego, USA

Abstract

Functional magnetic resonance imaging (fMRI) is a methodology for detecting dynamic patterns of activity in the working human brain. Although the initial discoveries that led to fMRI are only about 20 years old, this new field has revolutionized the study of brain function. The ability to detect changes in brain activity has a biophysical basis in the magnetic properties of deoxyhemoglobin, and a physiological basis in the way blood flow increases more than oxygen metabolism when local neural activity increases. These effects translate to a subtle increase in the local magnetic resonance signal, the blood oxygenation level dependent (BOLD) effect, when neural activity increases. With current techniques, this pattern of activation can be measured with resolution approaching 1 mm^3 spatially and 1 s temporally. This review focuses on the physical basis of the BOLD effect, the imaging methods used to measure it, the possible origins of the physiological effects that produce a mismatch of blood flow and oxygen metabolism during neural activation, and the mathematical models that have been developed to understand the measured signals. An overarching theme is the growing field of quantitative fMRI, in which other MRI methods are combined with BOLD methods and analyzed within a theoretical modeling framework to derive quantitative estimates of oxygen metabolism and other physiological variables. That goal is the current challenge for fMRI: to move fMRI from a mapping tool to a quantitative probe of brain physiology.

1. Introduction

Functional magnetic resonance imaging (fMRI) makes possible an experimental window to observe the working human brain. Figure 1 shows the measured responses in the motor area of the human brain based on MR signals sensitive to blood flow and blood oxygenation when subjects tap their fingers for 2 s. Even a brief stimulus elicits a strong blood flow change that translates to a weak *blood oxygenation level dependent* (BOLD) signal change. Unlike x-ray and nuclear medicine methods for measuring brain function, these fMRI measurements are completely noninvasive, requiring no injections of contrast agents or radioactive isotopes. Remarkably, this ability to probe functional changes within the intact brain is just based on physical principles of nuclear magnetic resonance (NMR) and the intrinsic effects of blood oxygenation on the MR signal due to the magnetic properties of deoxyhemoglobin.

The idea that changes in blood oxygenation could drive measurable signal changes in brain MR images was introduced by Ogawa and colleagues in 1990 and called BOLD contrast [1]. In a rat model they showed that venous vessels, and importantly the tissue near the vessels, had a low signal with an MRI technique that is sensitive to the local magnetic field heterogeneity [1]. When the animal breathed a gas mixture containing 10% CO_2, there was much less signal loss near the vessels. The physiological effect in this experiment is that breathing CO_2 dramatically increases brain blood flow, and at high levels reduces oxygen metabolism. The key effect was that the oxygen extraction fraction (OEF) in the brain—the fraction of O_2 carried by an element of blood that is removed in passing through the capillary bed—was reduced by breathing CO_2. The venous blood was thus more oxygenated, and the total amount of deoxyhemoglobin was reduced. In this initial demonstration the change in blood oxygenation was produced by an external agent (breathing CO_2). Ogawa *et al* [2] and Kwong *et al*, working independently [3], showed that *intrinsic* changes in blood oxygenation happen in normal physiology associated with changes in neural activity, marking the birth of fMRI as a tool for investigating patterns of activity in the brain. Interesting perspectives on the early development can be found in a recent issue of *Neuroimage* commemorating 20 years of fMRI [4–10].

The BOLD effect related to neural activity arises because of two distinct phenomena. The first is that when hemoglobin, the molecule in blood that carries oxygen, loses that oxygen to become deoxyhemoglobin, the magnetic properties change in a subtle way: deoxyhemoglobin is paramagnetic, and alters the magnetic susceptibility of blood [11, 12]. The difference in susceptibility between blood vessels and the surrounding tissue creates local magnetic field distortions that decrease the net MR signal. In the brain a typical OEF is ~40%, and in a 3 T magnetic field this level of deoxyhemoglobin in the veins and capillaries is sufficient to reduce the MR signal in the brain by ~10% in the baseline state compared with what it would be if no deoxyhemoglobin was present.

In short, the MR signal is sensitive to the OEF. This phenomenon alone, while interesting from a biophysical point of view, would not necessarily produce a useful basis for an experimental technique, because it is not obvious that the OEF would change with physiological activity. For example, matched fractional increases in cerebral blood flow (CBF) and cerebral metabolic rate of oxygen ($CMRO_2$) would leave the OEF unchanged. This biophysical effect becomes very useful, though, when combined with an unexpected physiological phenomenon: when an area of brain is activated, the blood flow increases much more than the oxygen metabolic rate [13]. This leads to a reduction in the OEF, a seemingly paradoxical scenario in which the venous blood is more oxygenated—despite the increase in oxygen metabolic rate—because the blood flow has increased more. Taken together, these two phenomena produce the BOLD effect, a local increase in the MR signal due to a reduction in the OEF during increased neural activity. Functional MRI based on the detection of BOLD signal changes has become the leading tool for imaging the working human brain. Nevertheless, a quantitative physiological interpretation of exactly what is being measured with BOLD-fMRI is complicated by the complexity of the signal, as discussed below.

To put fMRI in perspective in the broader context of other techniques for measuring brain function, it is useful to think of functional neuroimaging as having two main branches, depending on whether electromagnetic signals directly related to neural signaling are measured, or whether associated physiological changes (blood flow or metabolism) are measured. Although a primary goal of functional neuroimaging is to measure the electrical activity underlying neuronal signaling, localizing that activity with high spatial resolution is difficult without placing electrodes directly in the brain. Fluctuating electric potentials at the scalp and magnetic fields measured near the head provide information on electric currents within the brain, and from these data the location of sources of activity can be estimated with electroencephalography (EEG) or magnetoencephalography (MEG) methods. The alternate approach is to measure the changes in blood flow and metabolic activity that accompany neural activity changes, and this approach makes a more precise localization of the activity possible. Positron emission tomography (PET) methods were a critical advance in spatial localization, using radioactive tracers to measure blood flow, glucose metabolism and oxygen metabolism. These studies amply demonstrated the localized nature of blood flow and metabolism changes, suggesting a relatively tight linkage between neural activity, energy metabolism and blood flow. Functional MRI has better spatial and temporal resolution than PET methods, avoids the risks of ionizing radiation, and has largely replaced PET for functional neuroimaging research studies. Nevertheless, PET methods still play an important role for receptor studies and clinical applications.

A recurring theme in this review is that the BOLD signal is a complicated function of the underlying physiological changes, depending on the balance of changes in blood flow and oxygen metabolism. That is, the physiological changes measured with PET are one step away from the neuronal activity itself, but at least they clearly measure the defined physiological variables (e.g. CBF). The BOLD effect is not even a clean reflection of those physiological variables because it is primarily driven by the change in local deoxyhemoglobin concentration, which depends on the combined changes of CBF, $CMRO_2$ and the cerebral blood volume (CBV). The problem is that increased neural activity tends to increase each of these physiological variables, but these changes have conflicting effects on the BOLD response. Increased CBF tends to wash out the deoxyhemoglobin, while increased $CMRO_2$ increases local production of deoxyhemoglobin. Increased *venous* CBV increases the total deoxyhemoglobin content, partially offsetting the effects of the OEF change. In contrast, though, increased *arterial* CBV may increase the measured signal through a volume exchange effect as increased CBV pushes out extravascular fluid. Arterial blood typically generates a larger MR signal than tissue, and so this volume exchange leads to a positive signal change, added to the oxygenation-dependent change associated with deoxyhemoglobin changes. In short, the BOLD signal is a complex phenomenon, and much of this review is devoted to understanding that complexity and the difficulties involved in developing a quantitative interpretation of the BOLD signal. Nevertheless, the importance of a quantitative understanding is that it offers the potential for interpreting fMRI signals in terms of quantitative physiological variables, rather than simply a qualitative index of changing neural activity.

Functional MRI based on the BOLD effect is the most widely used method, but it is not the only MRI methodology that has been developed to be sensitive to local brain activation. Table 1 lists the primary physiological variables that change with neural activation and are thus potential targets for localizing and measuring activation. Arterial spin labeling (ASL) methods measure CBF [14], and figure 1(*a*) is an example of the CBF response to a brief stimulus. A number of approaches sensitive to CBV have been developed based on injection of an agent that remains in the blood vessels and alters the local MR signal in proportion to how much of the agent is present [15, 16]. More recently, a technique called *vascular space occupancy* (VASO) has been developed that is sensitive to changes in CBV and does not require injection of a contrast agent [17]. Magnetic resonance spectroscopy (MRS) methods have been developed to measure aspects of energy metabolism [18], although these methods have low spatial and temporal resolution and have not yet been widely adopted for human studies. Despite the clear importance of measuring $CMRO_2$ for understanding energy metabolism in the brain, it is a remarkably difficult physiological variable to measure [19]. Steady-state measurements are possible with PET, but require multiple tracers to account for confounding effects of CBF and CBV [20]. The basic sensitivity of the MR signal to deoxyhemoglobin clearly suggests the possibility that MR methods have the potential to provide measurements of $CMRO_2$, and current efforts in this direction are reviewed below.

This review focuses on the physical and physiological basis of the BOLD effect, with a view toward both understanding the basic mechanisms and guiding the interpretation of the BOLD signal in experimental applications. It is important to note that the underlying physiology is still poorly understood, and part of this review will necessarily be speculative. An overarching theme is that the complexity of the BOLD effect makes it difficult to interpret the magnitude of the BOLD response in a quantitative way. That is, a detected BOLD signal indicates that something is happening in that location, but if the BOLD signal magnitude differs between two groups (such as a healthy population and a disease population) it is difficult to interpret this experimental result in a quantitative way related to the underlying physiology. Nevertheless, a combination of techniques, particularly CBF measurements with ASL in conjunction with BOLD measurements, provides a much richer context for a quantitative interpretation of fMRI signals (e.g. see [21]).

The organization of this review is as follows. Section 2 gives an overview of how fMRI and related MR methods are currently being used to investigate brain function. Section 3 reviews the physiological basis of the BOLD effect, including current ideas about the links between the physiological variables accessible with MRI methods and the underlying neural activity. Section 4 reviews the physical basis of the BOLD effect, in particular the physics of MR signal decay in complex biological tissue. Section 5 considers the physiological information that can be derived potentially with MR methods, including the dynamics of different physiological variables. Section 6 is a summary and brief overview of the prospects for quantitative fMRI.

2. The fMRI experiment

2.1. The NMR signal

NMR is a highly developed field with many sophisticated methods for manipulating the magnetization associated with nuclear spins to yield informative signals. The basic method for generating the signal used in fMRI, though, is perhaps the simplest imaginable NMR signal. The central physical principles underlying NMR are the following:

1. *Equilibrium magnetization*. When placed in a magnetic field $\boldsymbol{B_0}$, the magnetic moments of nuclei with nonzero spin tend to weakly align with $\boldsymbol{B_0}$, creating a net macroscopic magnetization $\boldsymbol{M_0}$. Functional MRI manipulates the magnetization due to hydrogen nuclei (protons), and the hydrogen nuclei in the brain are overwhelmingly in water molecules (a proton concentration of about 78M, compared with mM concentrations of most other metabolites).

2. *Precession*. If the magnetization $\boldsymbol{M_0}$ is tipped away from alignment with $\boldsymbol{B_0}$, it will precess around the $\boldsymbol{B_0}$ axis with angular frequency $\omega_0 = \gamma B_0$, where γ (*gyromagnetic ratio*) is a constant for any given nucleus. For protons $\gamma = 2.675 \times 10^8$ rad T^{-1}, and a typical magnetic field for fMRI is 3 Tesla (T), so the precession frequency ν_0 ($=\omega_0/2\pi$) is approximately 128 MHz. After tipping the magnetization away from $\boldsymbol{B_0}$, the net magnetization vector can be described as two components: the remaining *longitudinal magnetization* along the $\boldsymbol{B_0}$ axis, and the rotating *transverse magnetization* perpendicular to $\boldsymbol{B_0}$. The rotating component generates an oscillating magnetic field that induces a current in a nearby coil, creating the basic measured NMR signal.

3. *Relaxation*. Over time, the transverse magnetization decays exponentially to zero with a time constant T_2, and the longitudinal magnetization recovers exponentially toward its equilibrium value M_0 with a time constant T_1. In gray matter in the human brain at a field strength of 3 T, T_1 ~1.0 s and T_2 ~ 0.1 s.

The NMR experiment begins with the nuclear magnetization aligned with $\boldsymbol{B_0}$. The central experimental manipulation of the magnetization is the application of a radio frequency (RF) transverse magnetic field at the resonant frequency ω_0, which has the effect of tipping the magnetization away from the $\boldsymbol{B_0}$ axis. The degree of tipping depends on the magnitude and duration of the *RF pulse* and is described by the *flip angle* (e.g. a 90° flip angle tips the magnetization completely from the longitudinal axis into the transverse plane). The newly created transverse component then precesses at frequency ω_0. This precessing magnetization creates a time-varying magnetic flux that induces a voltage oscillation in a nearby detector coil. The detected signal is an oscillation at frequency ω_0 that decays over time due to transverse relaxation, called a *free induction decay* (FID) (figure 2(*a*)). In practice, the FID decays faster than would be expected for the T_2 of the sample due to magnetic field inhomogeneities, and the decay constant is described as T_2^* (with $T_2^* < T_2$). One component of those additional field offsets is the field distortions around blood vessels containing deoxyhemoglobin, so the primary NMR effect of blood oxygenation is that it affects T_2^* (although there are important subtleties in this idea discussed below).

The basic distinction between T_2 and T_2^* is that the added field offsets that contribute to T_2^* signal decay are potentially reversible (with important exceptions discussed in section 4). In general, signal decay is a dephasing process due to the fact that the resolved signal (e.g. the net signal from a resolved volume element, or *voxel*, in a macroscopic imaging experiment) adds signals with different phases. These different phases result from precession in slightly different fields, and the key difference is whether these field offsets are varying randomly in time or are fixed in time. The source of T_2 decay is randomly varying fields, and for a hydrogen nucleus in a water molecule there is a significant contribution from the magnetic field of the other hydrogen nucleus in the molecule as the molecule tumbles. These random field fluctuations, different for each molecule, lead to a random walk of the phase and a resulting partial cancellation of the net signal. The effect of these random fluctuations is irreversible. In contrast, if two nuclei are located in different but constant fields, they will steadily get out of phase with each other, but this dephasing is reversible. Applying a 180° RF inversion pulse at a time TE/2 flips the magnetization in the transverse plane like a pancake, effectively reversing the sign of the accumulated phases. Continued phase evolution of each spin then unwinds the net phase accumulation acquired before the 180° RF pulse, so that at time TE (the *echo time*) the two signals are back in phase creating a *spin echo* (SE) (figure 2(*b*)). After TE the relative phases continue to evolve but another 180° pulse can bring them back together to form another echo. This SE process reverses the effects of static field offsets, but not the random field fluctuations, so with each SE the peak signal is reduced by T_2 decay.

In the field of MRI the usual terminology is that images in which the acquired signal is a spin echo are called SE images, while images in which the acquired signal is an FID are called *gradient echo* (GE or GRE) images. In both cases the delay after the initial excitation RF pulse is called TE, and TE controls the amount of signal decay that is allowed to happen before the signal is measured. The term ‘gradient echo’ is somewhat unfortunate here, because it tends to confuse two concepts. A GE occurs when an external linear gradient of the magnetic field is turned on as a short pulse, initially creating phase dispersion and signal loss, and a second gradient pulse with opposite sign is then applied to reverse that signal loss, creating an echo. Gradient echoes are essentially a technical trick used in imaging pulse sequences (including SE imaging), and do not have anything to do with intrinsic signal decay mechanisms. Spin echoes, though, are directly related to these intrinsic decay mechanisms. So even though a more logical terminology would be SE and FID imaging, we must work with SE and GE imaging, and for both methods the time delay between excitation and the center of data collection is called TE.

2.2. Image acquisition

While the NMR pulse sequence used in fMRI is among the simplest, the imaging methods are among the most sophisticated. A key technological innovation that made fMRI possible was the development of single-shot imaging methods that allowed collection of all the data needed for an image within a few tens of milliseconds after a single RF excitation. Roughly speaking, an MR image is a snapshot of the transverse magnetization at a particular time TE after the RF pulse that initially created the transverse magnetization. The essential physical relationship that makes imaging possible is the enormous range (seven orders of magnitude)

between the lifetime of the transverse magnetization ($T_2^* \sim 50\text{ms}$) and the period of the precession frequency ($2\pi/\omega_0 \sim 0.008$ μs). Because of this, quite small differences in frequency (less than one part per million) are readily detectable. The approach for imaging is to measure the NMR signal in the presence of rapidly adjustable linear magnetic field gradients that create small well-controlled field variations. The basic idea is that when a linear magnetic field gradient G in the x-direction is applied to a sample, the local resonant frequency becomes directly proportional to the position of the spins along the x-axis ($\omega(x) = \omega_0 + \gamma G x$). At time t, the local phase of the precessing transverse magnetization at x (the local signal) varies linearly with x. The net measured signal $S(t)$ is the phase-dependent sum of the signals from all positions. The key is that this process, implemented physically here, is identical to the mathematical process of calculating the Fourier transform (FT) of a function: a function $f(x)$ is multiplied by a cosine or sine function in kx and integrated to find the FT $F(k)$. The spatial frequency k is inversely proportional to the spatial wavelength. In the imaging process, the result of applying the gradient field is that the signal arising from each location is multiplied by a cosine or sine in phase $\varphi = \gamma G x t$, so identifying $k(t) = \gamma G t$ gives a direct relationship with the FT. In short, if $I(x)$ is the image of the spatial distribution of the MR signal, in the presence of a linear gradient the net measured signal over time $S(t)$ is directly proportional to the FT of $I(x)$, and image reconstruction is simply taking the inverse FT of $S(t)$.

This remarkable relationship between the net signal over time and the spatial FT of the distribution of transverse magnetization extends to two and three dimensions, in which field gradients in multiple directions are varied over time (figure 3). For example, a two-dimensional (2D) image $I(x, y)$ has a corresponding 2D k-space in the Fourier domain, and the net signal traces out a trajectory in k-space as the gradients are modulated. Specifically, by measuring the net signal over time while manipulating linear field gradients in x and y, $G_x(t)$ and $G_y(t)$, a range of k-space is mapped, and this range directly determines the resolution of the image. Higher resolution requires mapping k-space out to higher values of k (shorter spatial wavelength). Specifically, the *spatial resolution* of the image is the distance between two points in space for which the phase difference of their signals is 180° for the highest value of k measured.

These ideas illustrate the fundamental flexibility of MRI. Imaging can be based on any pattern of gradients that causes the net signal to follow a trajectory in k-space that traces out a sufficiently large block of k-space for the desired image spatial resolution. The most commonly used single-shot technique is *echo planar imaging* (EPI), in which the trajectory in k-space is a back and forth raster pattern (figures 3(*a*) and (*b*)). Specifically, the trajectory in k-space moves along a line in the $+k_x$ direction at fixed k_y (G_x = constant,$G_y = 0$), then a brief pulse of G_y shifts the trajectory to a new k_y level and a constant G_x with reversed sign moves the trajectory in the $-k_x$ direction at the new value of k_y. A typical acquisition uses 64 such scanning lines in k-space. Note that the amplitude of each point in k-space is the amplitude of a particular sinusoidal pattern across the image, i.e. one Fourier component (figures 3(*c*) and (*d*)).

This basic imaging scheme also highlights the technical hurdles involved. Time is limited for scanning through k-space, because for a single-shot image the entire sampling must be

done under the overarching window of the exponential decay of the MR signal (i.e. an interval of duration about T_2^*). Thinking of $k(t) = \gamma Gt$ for a constant gradient, the 'speed' at which the trajectory moves through *k*-space (i.e. Δk in a given time interval) is directly proportional to the magnitude of the gradient *G*, so stronger gradients mean that *k*-space can be sampled faster. But the back and forth EPI trajectory also means that this gradient must be rapidly reversed, and the time required to do this is determined by the *slew rate* of the gradient hardware (how quickly the sign of the gradient can be reversed). These two hardware characteristics, gradient strength and slew rate, determine the performance of the gradient system. Interestingly, current limits on MRI systems are not set by hardware limitations, but by the need to limit slew rates because the associated d*B*/d*t* can surpass nerve stimulation thresholds. An alternative to the back and forth trajectory through *k*-space of an EPI acquisition is a spiral acquisition trajectory created by oscillating both *x* and *y* gradients. A spiral trajectory provides an efficient balance of the limiting demands of gradient strength and slew rate in covering *k*-space [22].

The development of EPI was crucial for fMRI. Although it originated in the pioneering MRI work by Peter Mansfield [23], it did not begin to be available on human imaging systems until the early 1990s, just in time to be exploited for imaging the BOLD effect. The key advantage of EPI is that it is fast and also has a high signal-to-noise ratio (SNR). The latter effect may not be obvious, but stems from the low spatial resolution of EPI, which is limited by the range of *k*-space that can be covered before the signal decays away. The SNR is proportional to the volume of a resolved tissue element (voxel), and for EPI the SNR can be on the order of 100 : 1. This makes this imaging technique well-suited to measure the dynamic, weak signal changes associated with the BOLD effect.

2.3. Limits of spatial and temporal resolution for human brain imaging

The human cerebral cortex is a convoluted sheet with an area of ~2000 cm^2, a thickness of ~2.5mm and ~10^5 neurons mm^{-2} (i.e. in a column through the cortex with an area of 1 mm^2) [24]. A whole brain image acquisition requires coverage of a three-dimensional (3D) rectangular volume of about 200 × 180 × 180 mm^3. For typical whole brain fMRI studies the spatial resolution is usually about 3 × 3 × 3 mm^3 with a temporal resolution of about 3 s. (These numbers can be improved if one abandons whole-brain coverage and focuses on a smaller volume of brain, but for many studies whole brain coverage is the goal.) The standard acquisition is a series of 2D images, in which the effect of the RF excitation pulse is limited to one slice, and after collecting the *k*-space data for that slice an RF excitation pulse is applied to the next slice, etc. After 3 s, the first slice is acquired again. An alternative approach is a full 3D acquisition in which gradients in all three spatial directions are rapidly modulated to create a signal trajectory through a 3D *k*-space. Currently, most fMRI studies use the 2D multi-slice approach.

However, the methodology for image acquisition continues to develop in sophistication, leading to demonstrations of dramatic improvements in spatial and temporal resolution in recent years. The central challenge is that these two goals conflict with each other: better spatial resolution means that more of *k*-space must be sampled, and that takes more time. The central technological innovation that has driven improvement in spatial and temporal

resolution is the introduction of arrays of multiple small receive coils (*multi-channel coils*) instead of one large RF coil, often described as *parallel imaging* because the NMR signals are simultaneously measured with multiple coils and receivers. The key improvement here is that each of the smaller coils has a limited volume of sensitivity for detecting the NMR signal, so each has a unique spatial sensitivity pattern. This added spatial information from the coil locations means that it is not necessary to cover all of *k*-space to reconstruct an image with high spatial resolution. By limiting the points in *k*-space that must be sampled for each image, the imaging time can be reduced several fold without sacrificing spatial resolution.

In addition, with a *multi-band* approach, several widely separated 2D slices can be acquired at the same time, and the signals separated using the spatial information from the different coil sensitivity patterns. By combining the parallel imaging and multi-band approaches, Moeller *et al* demonstrated whole brain fMRI acquisitions with spatial resolution of $1 \times 1 \times 2$ mm^3 and temporal resolution of 1.5 s [25]. Feinberg *et al* introduced a multiplexed method that adds a third approach to these two in which a few adjacent slices are excited with slight delays, so that the data from all three slices are measured in a longer data readout window [26]. With this multiplexed approach they demonstrated acquisition of fMRI data sets with 2mm isotropic resolution in less than 1 s. This method is currently being used in a large-scale project to map the patterns of connectivity in the human brain (the Human Connectome Project) [27].

The basic goal of increased spatial resolution with decreased acquisition time has a cost in reduced SNR. For this reason, ultrafast imaging improves at higher magnetic field strengths, where the intrinsic strength of the NMR signal increases because the equilibrium magnetization increases in proportion to B_0. The current workhorse for human fMRI is the 3 T system, but there are currently > 30 7 T systems around the world [28], and a few 9.4 T systems. At the higher fields, several studies have pushed the limits of spatial resolution to resolve functional activity on scales around or below 1mm in humans. Recently, Olman *et al* demonstrated changes in the distribution of the BOLD response across the different layers of the visual cortex with 0.7mm isotropic resolution at 7 T [29]. In addition to layer-specific activity, the visual cortex is organized in columns on the order of 0.5mm across, with similar neuronal responses within a column (see [30] for a cautionary discussion about the heterogeneity of the cortex and the different uses of the term ‘column’ in neuroscience). A number of studies have used high-field fMRI to probe functional activity at the columnar level [31, 32].

2.4. fMRI experimental designs

Just as the technology for fMRI studies has continued to improve, the design of experiments has evolved in sophistication as well. The classic fMRI experiment is a simple block design, such as an alternation between 30 s of performing a task and 30 s of baseline, with four blocks in an experimental run that lasts 4 min. For example, a simple experiment often repeated in the early days of fMRI was alternation between finger tapping and rest (as in figure 1). During this time dynamic images are collected, typically with 2D-EPI images moving sequentially through different slices to cover the brain (figure 4). Current scanners

can support about 20 EPI acquisitions per second, so that the brain can be covered with 3 mm thick 2D images (a full brain volume acquisition) in about 3 s, so that dynamic images on any single slice are about 3 s apart. (The newer pulse sequence designs described in the last section have not yet become standard on 3 T systems used for fMRI.)

The basic analysis then involves correlating the measured time series for each voxel with a reference model function defined by the stimulus pattern (e.g. the 30 s on/off blocks) [33]. Voxels that have a significant correlation are identified as activated by the chosen task. The model function is essentially a guess as to what the fMRI signal from a brain region responding to the task would look like. Because the *hemodynamic response function* (HRF) is slow, taking several seconds to rise to a new level with a block stimulus, the model function is usually taken as a delayed and rounded version of the stimulus pattern. The method is referred to as a *general linear model* (GLM) approach [34, 35], based on the idea that a voxel's time course is described as a linear combination of a scaled version of the model function plus random noise. In addition, other sources of systematic signal variation, such as scanner drift and physiological noise due to breathing and heart pulsations, can be included as additional nuisance regressors in the model. This bare-bone description hides much of the sophistication of the statistical techniques that have been developed for analyzing fMRI data, and the field of statistical analysis has grown enormously since the early days of simple block design experiments [36].

An important innovation was the introduction of event-related designs (for a review, see [37–40). Rather than grouping stimuli in a block, they are presented randomly, making it possible to ask questions that cannot be easily answered with block designs. For example, in a memory task it would be useful to present items that had been previously studied and novel items to measure differences in the patterns of activation. This is difficult with a block design because grouping the stimuli prevents analyzing the responses individually. With an event-related design one can use different model functions for novel and studied stimuli based on the known pattern of application of these stimuli. In addition, though, by measuring responses from the subjects on whether they remember each stimulus, one could also separate the previously studied stimuli into those that were remembered and those that were not, and develop appropriate model functions for each, and thus identify brain areas related to successful recall. In addition, with an event-related design it is also possible to estimate the HRF from the data, provided the data include samples in a range of delays after each stimulus type was applied [41].

While block designs and event-related designs are widely used for determining response patterns to specific stimuli, in recent years a great deal of research has focused on using fMRI to explore functional connections between brain regions in the absence of any applied stimuli [42, 43]. For example, an early discovery was that even when the subject was not performing a motor task, the BOLD signal fluctuations in the motor areas of the left and right hemispheres were correlated [44]. This approach has expanded enormously, and a number of recurring *resting state networks* (RSNs) have been identified as brain regions with coherent BOLD fluctuations [45, 46]. Newer methods are being developed to look at the temporal evolution of these connectivity patterns [47–49].

Brain decoding is another approach looking at patterns of activity across the brain rather than correlating each voxel with a reference model function. The essential goal is to use a training set of data showing how an individual subject's brain responds to particular stimuli or conditions, and then use multivariate analysis methods to classify the overall pattern of activity in terms of those stimuli or conditions [50, 51]. These methods have shown a remarkable ability to identify what a person is seeing or remembering, just from patterns of fMRI activation, leading to popular accounts describing these methods as 'mind reading'. Overall, these studies nicely demonstrate that there is a great deal of information about the ongoing neural activity in the brain reflected in the patterns of transient blood flow changes measured with fMRI.

2.5. Diffusion tensor imaging for mapping anatomical connections

In addition to BOLD-fMRI activation experiments, a number of studies are integrating this functional information with anatomical connectivity information based on *diffusion tensor imaging* (DTI) [52]. Classically, NMR has been one of the most sensitive ways to measure the diffusion of water [53]. In the brain, water molecules randomly migrate over time due to thermal motions with a diffusion constant $D \sim 1\ \mu m^2\ ms^{-1}$ ($10^{-5}\ cm^2\ s^{-1}$). In one dimension, molecules starting at the same location will spread into a Gaussian distribution of displacements with variance $\sigma^2 = 2DT$ at time T, so that in the brain molecules randomly move on the order of 12 μm in 100 ms. Although this displacement is much smaller than an imaging voxel, the effects of diffusion can be encoded in the NMR signal in a detectable way that makes it possible to measure D.

In its simplest form, the method involves adding a bipolar gradient pulse to the FID pulse sequence (figure 5(*a*)). After excitation, imagine the local magnetization vectors at different locations in x precessing uniformly. A strong linear x-gradient pulse with amplitude G and short duration δt is then applied to create a position-dependent phase variation $\varphi(x) = \gamma G x \delta t$. After a delay T the same gradient pulse, except with opposite sign, is applied. If the water molecules have not moved during the delay T, the phase offsets due to the two lobes of the bipolar gradient pulse will be precisely opposite in sign. The local precessing magnetization vectors are then all back in phase, so that the bipolar gradient pulse would have no effect on the net signal. However, if the water molecules move randomly in x due to diffusion during the interval T, the phase offsets due to the gradient pulses will not cancel. The net phase difference after the bipolar gradient pulse is proportional to Δx, the distance moved during T. For diffusion, Δx has a mean of zero and a variance proportional to D. The net signal is calculated by multiplying this Gaussian distribution of phases by $\cos\varphi$ and integrating. The result is that the net signal is attenuated by the bipolar gradient by a factor $\exp(-bD)$, where b is a lumped factor depending on the amplitude and timing of the bipolar gradient pulse (figures 5(*a*) and (*b*)). Increasing b by increasing the strength or duration of the gradient pulses improves sensitivity for measuring diffusion effects, and this is often a motivation for the development of stronger gradient hardware on MRI systems. Figure 5(*c*) illustrates the effect of diffusion on the acquired image, along with a map of the estimated value of D, often referred to as the *apparent diffusion coefficient* (ADC).

This basic experiment measures attenuation due to diffusion along a particular spatial axis, the direction of the bipolar gradient pulse. It can be repeated many times to map out displacements due to diffusion along many directions. In the Gaussian diffusion model particles that start at the same location are displaced in an ellipsoidal pattern over time, with each of the three *principal axes* of the ellipsoid having a different value of *D*. The MR measurement of diffusion, though, is sensitive just to displacements along a single axis, and in general this axis will not correspond to one of the principal axes. That is, the 3D ellipsoidal pattern of displacements must be projected onto the single axis defined by the diffusion gradient direction, and it is the displacements along that axis that determine the attenuation of the MR signal. The measured value of *D* along an arbitrary axis is determined by the *diffusion tensor*, a 3 × 3 symmetric matrix with six independent components [54, 55]. Physically, there are six independent numbers because a description of the diffusion ellipsoid requires two angles to specify the orientation of the first principal axis, an additional angle to specify the orientation of the second principal axis relative to the first, and three values of *D* for the three principal axes. By measuring diffusion along at least six axes the diffusion tensor can be calculated, and from the tensor the principal axes of diffusion are derived. When the diffusion tensor is expressed in a coordinate system defined by the principal axes, the diagonal elements are the values of *D* for the principal axes and the off-diagonal elements are zero.

The first principal component is the axis along which diffusion is highest. For gray matter in the brain, diffusion is reasonably isotropic. Importantly, though, for white matter diffusion is highly anisotropic. White matter contains nerve bundles of axons, the long range wiring of the brain connecting neurons in different brain regions. In these white matter fiber tracts diffusion of water is much higher along the fiber than perpendicular to the fiber, so identifying the first principal component of the diffusion tensor identifies the local orientation of the fibers. By connecting these local orientation vectors the fiber tracts can be traced out, providing measures of anatomical connectivity between one brain region and another [56, 57]. For a complex structure with multiple crossing fibers, the classic diffusion tensor cannot provide a complete description of local diffusion. Expanding the number of directions along which diffusion is measured provides a more complete characterization of local diffusion [58–60]. The field of fiber tract mapping with DTI has grown enormously over the last decade, providing highly detailed information on brain anatomical connections, although more work is still needed to validate these maps [61].

2.6. ASL for measuring CBF

While the BOLD signal provides a reasonably sensitive measure for detecting changes in brain activity, it is difficult to interpret quantitatively for two reasons. First, the BOLD signal is only sensitive to a change between two states during the experiment (e.g. a baseline state and a task state). For this reason, chronic changes in the baseline state, as might be the case in disease or in an aging population, cannot be detected unless the baseline state change alters the acute response in the fMRI experiment [62]. Second, the BOLD effect is a complex reflection of the underlying physiological change. Because it primarily depends on the change in deoxyhemoglobin, the blood flow and oxygen metabolism changes with

activation drive the BOLD signal in opposite directions, so that the BOLD signal magnitude strongly depends on the exact balance of the two physiological changes.

An alternative methodology, designed to overcome these limitations by specifically measuring CBF, is called *arterial spin labeling* (ASL) [14, 63, 64]. CBF is defined as the rate of delivery of arterial blood to a tissue element. For the simplest ASL experiment the goal is to create two images in which the magnetization of the delivered arterial blood has been manipulated, but the static signal from the rest of the tissue remains the same (figure 6). Subtraction of these images then yields an image just of the delivered arterial blood, and so is proportional to local CBF. For example, applying a 180° inversion pulse to a band below the slice of interest will invert the magnetization of blood in the arteries, and after a delay TI (*inversion time*) some of the labeled blood is delivered to the imaging slice where it adds to the net signal with a negative sign due to the inversion. In the second experiment, the arterial magnetization is not inverted, and after the same delay to allow for delivery to the imaging slice it adds to the net signal with a positive sign. If this is done carefully so that the magnetization of the rest of the tissue is precisely the same for the two images, subtraction will leave just a signal proportional to the delivered arterial blood during the delay TI.

In practice, ASL techniques have evolved in sophistication to deal with a number of experimental issues, including relaxation of the labeled magnetization, transit delays from the tagging region to the imaging slice, and accurate balancing of the labeling and control pulses to deal with inversion efficiency and off-resonance excitation effects, among others [65–70]. With current ASL methods it is possible to accurately measure CBF, although the sensitivity and spatial and temporal resolution are typically worse than what is currently available with BOLD imaging (compare the noise levels in the two measured responses in figure 1). For this reason, ASL has not replaced BOLD imaging for activation studies, but provides additional quantitative physiological information. By providing a quantitative measurement of a specific, well-defined physiological variable it is more readily interpretable than the BOLD response, and it also provides a measurement of CBF in the baseline state, overcoming some of the difficulties of interpreting BOLD responses in disease. As discussed below, the ability to measure both CBF and BOLD responses opens possibilities for a much more quantitative assessment of brain function, including estimation of oxygen metabolism changes.

3. The physiological basis of fMRI

3.1. Hemoglobin and blood oxygenation

The primary energy source for brain neural activity is the oxidative metabolism of glucose to carbon dioxide (CO_2) and water, with one CO_2 molecule produced for each O_2 molecule metabolized. Local brain metabolism requires constant delivery of oxygen and constant clearance of CO_2 by blood flow. The essential problem in transporting O_2 through the body is that it has a low solubility in water. Carbon dioxide, in contrast, readily dissolves by chemically combining with water to form bicarbonate ions. If O_2 and CO_2 as gases are maintained at the same partial pressure above a surface of water, the concentration of dissolved CO_2 in the water is about 30 times higher than that of O_2. The clearance of CO_2 by blood flow is then relatively simple due to this high CO_2-carrying capacity, and the

delivery of O_2 is the difficult task. Evolution has solved this problem with carrier molecules that readily bind oxygen in the lungs and then release it in the capillary. In mammals, this molecule is the hemoglobin (Hb) contained in the red blood cells, and it increases the O_2-carrying capacity of blood by about a factor of 30–50. The O_2–Hb binding curve—the fractional saturation of Hb as a function of the plasma partial pressure of O_2—has a sigmoidal shape, as illustrated in figure 7. As long as pO_2 is above about 80 mmHg, the arterial Hb is nearly fully loaded with oxygen. The point at which the hemoglobin is half-saturated is called the p_{50} of the hemoglobin, about 27 Torr for human blood at body temperature. A number of factors, such as pH or temperature, change the p_{50} and shift the dissociation curve to the left or right [71].

Figure 7 compares typical atmospheric pO_2 values for sea level and the top of Mt Everest with approximate dissolved gas pO_2 levels in different vascular compartments and brain tissue to illustrate the gradients involved in O_2 transport. Although the dissolved gas component is relatively unimportant in terms of carriage of oxygen to the capillary bed, it is the key component for the actual transfer of O_2 from blood to tissue. When the blood reaches the capillaries (and to some degree the arterioles as well), the dissolved O_2 in plasma diffuses out of the vessel into the tissue. Because the O_2 bound to hemoglobin and the dissolved O_2 are in rapid equilibrium, O_2 is released from the hemoglobin and partly replenishes the dissolved gas in the plasma. As more oxygen leaves the blood, the plasma pO_2 largely follows the O_2 saturation curve. However, the CO_2 increasing in the blood also leads to a rightward shift of this curve, which complicates modeling of the gas exchange.

3.2. Blood flow and oxygen metabolism changes with brain activation

CBF is a somewhat subtle concept, in that it is not directly related to the motion of blood within the capillary network of an element of tissue. Instead, CBF is defined in terms of *delivery* of arterial blood to the tissue element, specifically the volume of arterial blood delivered to the tissue element in a unit time divided by the mass (or volume) of the tissue element. Classic units of CBF are ml/100 g per min, and a typical value for the human brain is 50 ml/100 g per min. For imaging experiments a more natural measure of the volume element is its volume, rather than its weight, and because the density of the brain is close to 1 g ml^{-1}, the CBF value in these units is then about the same. Note that CBF then essentially has units of inverse time (ml blood per ml tissue per min), and we can think of a basic value of CBF as 0.5 min^{-1}. In 2 min the volume of arterial blood delivered to a tissue element is about equal to the volume of the tissue element.

The $CMRO_2$ of a tissue element is defined as the moles of O_2 consumed within the element per minute divided by the volume of the tissue element, and a typical value in the human brain is 1.6 μmoles ml^{-1} min^{-1}. Note that this can be written as a concentration divided by time, 1.6mM min^{-1}. OEF is the fraction of the O_2 molecules delivered to the capillary bed that are extracted and metabolized. The basic relationship between CBF (F), $CMRO_2$ (J_{O_2}) and OEF (E) is essentially just mass balance:

$$J_{O_2} = EF[O_2]_a \quad (1)$$

where $[O_2]_a$ is the concentration of O_2 in arterial blood, with a typical value of about 8mM. The product of CBF and $[O_2]_a$ is simply the rate of delivery of O_2 to the capillary bed, and the fraction metabolized is then $CMRO_2$. In a healthy subject at sea level the hemoglobin of arterial blood is nearly saturated (~98%), and most of the arterial O_2 is bound to hemoglobin (only a few percent of the total is present as dissolved gas in the plasma). For this reason, $[O_2]_a$ primarily reflects the concentration of hemoglobin in arterial blood (with four O_2 molecules bound to each hemoglobin). This means that the hematocrit, the volume of blood occupied by red cells containing hemoglobin, affects $[O_2]_a$.

The physiological phenomenon at the heart of the BOLD effect is that the fractional increase in CBF with activation is about twice as large as the fractional increase in $CMRO_2$. From equation (1), this leads to a *decrease* in the OEF *E*, and a resulting increase in the MR signal. Why this seeming imbalance of CBF and $CMRO_2$ changes occurs is unknown, but a possible explanation for the function served by decreasing *E* with activation is that this preserves the oxygen concentration in the tissue (figure 7(*a*)) [19, 72]. Returning to the view of O_2 diffusing from plasma to tissue, the diffusion gradient from blood to tissue is proportional to the difference of the concentrations of dissolved O_2 in the blood and tissue spaces. For $CMRO_2$ to increase, the pO_2 gradient between blood and tissue must increase, which could happen in two ways: opening previously closed capillaries (capillary recruitment), so that the diffusion distance is reduced, or, increasing the pO_2 difference between capillaries and tissue. Current thinking is that capillary recruitment is a small effect, if it happens at all [73], so to increase the capillary/tissue pO_2 difference, either the tissue pO_2 must drop or the capillary pO_2 must rise. The key relationship for the latter possibility is that for capillary pO_2 to rise, the OEF must fall. A simple O_2 transport model suggests that CBF must increase 2–3 times more than $CMRO_2$ in order to preserve the tissue pO_2, consistent with the experimental observations [19] (figure 7(*b*)). Recent experimental evidence suggests that it may not be the mean tissue pO_2 that is important to preserve, but rather the pO_2 in regions with the lowest value [72]. In short, the evolutionary benefit of the mismatch between CBF and $CMRO_2$ changes with activation—the physiological effect that leads to the BOLD effect on the MR signal—may be a homeostatic mechanism to preserve tissue pO_2.

A possible reason why preserving tissue pO_2 is an important physiological function is that there is very little reserve of O_2 in the tissue to act as a buffer [19]. Because O_2 is so poorly soluble in water (solubility 0.0013mM/mmHg at body temperature), a typical tissue pO_2 of 25 mmHg corresponds to a concentration of only ~0.03mM. For a typical $CMRO_2$ of 1.6mM min^{-1}, if delivery of O_2 stopped the O_2 dissolved in tissue would be depleted in ~1 s. The blood within a tissue element provides somewhat more of a buffer. For a blood volume fraction of 4% and a rough average hemoglobin saturation of 70%, the total O_2 concentration in a volume of tissue approaches 0.3mM, and the depletion time rises to ~10 s. In short, the brain is critically dependent on a continuing supply of O_2 delivered by blood flow.

3.3. Energy costs of neural activity

The change in the OEF that drives the BOLD response depends on the balance of changes in CBF and $CMRO_2$, so it is important to consider two questions. (1) What are the energy costs of neural activity that would drive $CMRO_2$? (2) What are the mechanisms that drive the CBF change when neural activity changes? In line with the speculation above that the biologically useful function of the large CBF change is to maintain tissue pO_2, one could imagine a simple feedback system in which CBF is controlled by an oxygen sensor in tissue, so that increased energy metabolism associated with neural activity then drives the CBF change. Interestingly, though, a substantial body of evidence suggests that this is not the case. An appropriate oxygen sensor has not been found, and instead a number of mechanisms have been identified by which aspects of neural activity itself drive CBF. The current picture is that the acute changes in CBF are driven in a feedforward way by the neural activity, rather than the energy metabolism change [74]. This means that we must think of CBF and $CMRO_2$ as being driven in parallel by neural activity. While there is an overall pattern for the CBF change to be about twice as large as the $CMRO_2$ change, this ratio is not necessarily fixed. In this section and the next the questions of the energy costs and the CBF drivers are considered in turn.

The brain represents only about 2% of body weight, yet accounts for about 20% of the body's energy metabolism, and most of that energy cost is related to neuronal signaling. All cellular processes, including neural signaling, are constrained by the physical laws of thermodynamics: for any transformation of the system energy is conserved, and entropy increases or at best stays the same. The concept of *free energy change* (or Gibbs free energy change) combines the first and second laws of thermodynamics into a single useful relationship (see [75] for an excellent discussion of free energy in biological systems). For any transformation, such as a chemical reaction or movement of an ion across a membrane, there is an associated free energy change ΔG. The free energy is a measure of how far a system is from equilibrium, with a negative value of ΔG meaning that the transformation moves the system closer to equilibrium. Cellular work refers to processes that have a positive ΔG, moving a part of the system away from equilibrium, such as transport of an ion against its electrochemical gradient. A process with a positive ΔG can only occur if it is tightly coupled to another process with a more strongly negative ΔG, so that the net ΔG for the combined transformation is negative. That is, in order to move one system farther from equilibrium, that transformation must be coupled to another system that is moving closer to equilibrium.

For understanding the thermodynamic basis of neuronal signaling, the two key systems that are far from equilibrium are the adenosine triphosphate (ATP) system (a chemical reaction out of equilibrium) and the sodium ion (Na^+) gradient across the cell membrane (a diffusion gradient out of equilibrium). These two systems are the primary sources of negative ΔG to drive thermodynamically uphill reactions. The breakdown of ATP to adenosine diphosphate (ADP) and inorganic phosphate (Pi) is at equilibrium with a very low ratio of [ATP]/[ADP], yet in the body the ATP/ADP ratio is much higher, leading to a strong negative ΔG for the transformation ATP $\rightarrow$ ADP + Pi [76]. The Na^+ distribution across the cell membrane is also far from equilibrium, with a high extracellular concentration and a low intracellular

concentration. In addition, the intracellular space has a negative electric potential relative to the extracellular space, making the Na^+ gradient even farther from equilibrium. For this reason, a sodium ion moving from the extracellular to intracellular space also is associated with a negative ΔG, although not as strong as the negative ΔG for ATP. The potassium (K^+) distribution also is out of equilibrium, but with a higher concentration inside the cell than outside the cell it is closer to equilibrium than the Na^+ distribution.

Pumping a Na^+ ion out of the cell against its gradient—a strongly uphill thermodynamic process—can be done by coupling the transport to the more strongly downhill process of conversion of ATP to ADP. The enzyme that catalyzes this combined process is the *sodium/potassium pump*, which couples the uphill processes of moving sodium and potassium against their gradients to the downhill conversion of ATP to ADP. In this way we can think of the ATP/ADP system and the Na^+ distribution across the cell membrane as two batteries able to drive cellular work, with the ATP/ADP system able to recharge the Na^+ gradient through the sodium/potassium pump. This analogy with batteries of successively higher voltage extends to cellular signaling mechanisms as well [77]. Calcium ions (Ca^{2+}) also are far from equilibrium, with a higher concentration outside than inside the cell. The intracellular Ca^{2+} concentration often serves as the primary signal to initiate cellular activity. For example, the arrival of an action potential at a synapse with another neuron opens Ca^{2+} channels on the pre-synaptic side, so that Ca^{2+} flows into the cell (a downhill process). The rising intracellular Ca^{2+} concentration then triggers the release of neurotransmitter into the synaptic cleft. One way of restoring the initial Ca^{2+} gradient is by coupling movement of Ca^{2+} against its gradient to movement of Na^+ down its gradient [78], recharging the Ca^{2+} battery.

In neural signaling, the arrival of an excitatory action potential triggers a downhill cascade of events, including opening Ca^{2+} channels so that calcium flows into the presynaptic terminal, release of neurotransmitter into the synaptic cleft, binding of neurotransmitter to a receptor on the postsynaptic neuron, and opening of Na^+ channels with a rush of Na^+ ions down their gradient. In this way, the Na^+ gradient acts like an amplifier, with the opening of Na^+ channels as a switch that connects a strong battery to a circuit. Neural signaling itself is a downhill process because the system is maintained far from equilibrium, and the energy cost is in recovery, primarily pumping back the ions against their gradients (figure 8). Clearing neurotransmitter and resetting the pre-synaptic terminal for the next signal is a relatively small part of the energy cost compared with pumping Na^+, the amplifier signal, against its gradient. For this reason, the sodium/potassium pump consumes most of the ATP needed for recovery from excitatory neural signaling [79]. Attwell and co-workers developed a detailed energy budget for the primate brain based on the associated ATP costs, and concluded that about 74% of the energy cost was related to synaptic activity, and only about 20% to the generation of action potentials [74, 80]. The overall energy consumption is closely related to the spiking rate [81], but it is the integrative activity associated with a neuron receiving many synaptic inputs that is costly [82].

The energy cost of inhibitory neuronal signaling may be considerably lower. At an inhibitory cortical synapse the primary action is the opening of chloride (Cl^-) or potassium (K^+) channels. The chloride concentration ratio across the membrane is similar to that of

Na^+, but because of its negative charge it is approximately in equilibrium with the resting membrane potential. The potassium distribution is opposite to that of Na^+ (high intracellular concentration, low extracellular concentration), and would be in equilibrium with a more negative membrane potential. The membrane potential itself is a balance between the equilibrium potentials of the different ions weighted with the number of channels open for each ion (membrane permeability), and if the potential becomes sufficiently depolarized it fires an action potential. An excitatory signal opens Na^+ channels, depolarizing the membrane potential and moving the cell toward firing. Opening chloride channels tends to stabilize the resting membrane potential, while opening K^+ channels would hyperpolarize the neuron (i.e. move it farther from firing). Because the ions related to inhibitory signals are closer to equilibrium than Na^+, the associated currents and energy costs of pumping the ions back should be substantially less than for excitatory activity.

Neural activity is thus maintained by the consumption of ATP, and the ATP/ADP system is restored by the oxidative metabolism of glucose [83]. Glucose and O_2 are delivered by blood flow, and in the cytosol the first broad step of metabolism is glycolysis, the conversion of a glucose molecule to two molecules of pyruvate. This step is coupled to the net conversion of two ADP molecules to two molecules of ATP. The two pyruvate molecules then diffuse into the mitochondria, where they are broken down with a net consumption of 6 O_2 molecules, with production of 6 CO_2 molecules and ~32 ATP molecules. (Intermediate stores of free energy in this complex process include both chemical forms, such as the NADH/NAD^+ ratio, and an H^+ gradient across the inner membrane of the mitochondria.) The full picture of the dynamics of glucose metabolism in the brain is still somewhat unclear. With neural activation, glucose metabolism increases more than O_2 metabolism [84], suggesting that not all of the pyruvate produced by glycolysis is going into the mitochondria for oxidative metabolism. A leading theory is that astrocytes, non-neuronal glial cells that play a key role in recycling neurotransmitter at a synapse, preferentially use glycolysis to generate ATP [85]. The excess pyruvate is then converted to lactate and released for uptake by the neurons, where it is converted back to pyruvate and metabolized in the mitochondria (*lactate shuttle hypothesis*). Despite the excess glycolysis, the majority of the energy costs are met by oxidative metabolism [86].

In summary, the primary energy cost related to neuronal signaling is estimated to be due to excitatory synaptic activity, and oxidative metabolism of glucose provides most of the energy. We expect that the change in $CMRO_2$ in a brain region reflects the overall energy cost of the neural activity.

3.4. Neurovascular coupling

As noted above, current thinking is that the acute change in CBF associated with neural activity is driven by aspects of neural activity itself rather than by the energy metabolism change. There are also feedback mechanisms, undoubtedly, that serve to adjust CBF to the demands of energy metabolism, but these likely operate on a longer time scale. Numerous experiments have revealed particular pathways involved in the control of CBF [87–91], but how these pathways function in a coordinated dynamic network is still largely unknown. The following section briefly describes some of the current ideas.

Blood flow increases when the smooth muscle surrounding the arteries and arterioles relaxes, allowing the vessel to dilate and thus reducing the resistance. Tension created by the smooth muscle cells is primarily related to the cytosolic Ca^{2+} concentration, which depends on exchange between the cytosol and Ca^{2+} stores within the cell, and on the influx of Ca^{2+} from the extracellular space through voltage-sensitive Ca^{2+} channels, which tend to open as the cellular membrane depolarizes [92]. For this reason, cytosolic Ca^{2+} tends to follow the membrane potential, with a graded depolarization producing a graded increase in Ca^{2+} and a corresponding contraction of the smooth muscle. Because of this sensitivity to the membrane potential, a number of agents are thought to exert an effect on the arterial diameter by opening K^+ channels on the smooth muscle cell. Opening K^+ channels hyperpolarizes the cell, reducing cytosolic Ca^{2+} and relaxing the muscle. Potassium channels are remarkably diverse, making it possible for a number of vasoactive agents to modulate CBF by opening or closing potassium channels. Other mechanisms interfere with the way cytosolic Ca^{2+} couples to the enzymes that control muscle contraction. For example, nitric oxide (NO) initiates a chain of events leading to production of the intracellular messenger cyclic guanine monophosphate (cGMP), and the cGMP is thought to affect both K^+ channels and the sensitivity of the contractile mechanism to Ca^{2+}.

A wide range of vasoactive agents have been identified, and a partial list suggests the overall importance of controlling blood flow and the many mechanisms that have evolved to control it. Cations including H^+, K^+ and Ca^{2+} in the fluid space around the arteries dilate the vessels [93, 94]. The latter two suggest a possible connection with neural activity, while the former suggests a mechanism linking to energy metabolism through pH, as excess lactic acid dilates the vessel. Adenosine has a strong vasodilatory effect [87, 95, 96], and is interesting because of its intertwined roles in both neural activity and energy metabolism. As discussed earlier, the primary energy storage molecule is ATP. In addition to its role in energy metabolism, ATP also serves as a neurotransmitter/neuromodulator [97]. In the extracellular space, ATP is sequentially broken down to produce adenosine, which then has a potent effect by reducing neuronal excitability. Caffeine competes for adenosine receptors [98], and a number of studies have shown that caffeine reduces CBF [99], consistent with the idea of a reduction of the vasodilatory effect of adenosine. A number of studies have shown that CBF is modulated through an extended metabolic pathway related to arachidonic acid (AA) and its derivatives [88, 100], including prostaglandins formed from cyclooxygenase (COX) pathways. Finally, a number of direct neural pathways and associated neurotransmitters have been found to have vasoactive effects, with norepinephrin and neuropeptide Y producing vasoconstriction, and vasoactive intestinal peptide (VIP), acetylcholine, NO and other transmitters, producing vasodilation [89, 101].

Current thinking is that many of these processes are related to the activity of astrocytes [90, 91, 102–104]. Astrocytes contain receptors for numerous neurotransmitters, including glutamate, GABA, acetylcholine and adenosine, and activation of these receptors induces changes in cytosolic Ca^{2+}. With numerous processes contacting neuronal synapses, the astrocytes are well positioned to play a key role in recycling neurotransmitter, and also to monitor and integrate local neuronal activity. Additional processes, called *end-feet*, make contact with blood vessels, so that astrocytes create a bridge between neuronal activity and

blood flow. This close anatomical arrangement has long suggested an important functional arrangement, and in recent years the mechanisms by which changes in neural activity translate into changes in CBF have become clearer. Because of the close interactions between neurons, astrocytes and blood vessels, this combination is often referred to as the *neurovascular unit* [105, 106].

Many of these mechanisms are consistent with a basic picture in which excitatory neural activity increases CBF, consistent with the idea that this is also associated with the primary energy cost of neural signaling, and that inhibitory neural activity decreases CBF (vasoconstriction). However, there are a few interesting exceptions. As noted above, adenosine has an inhibitory effect on neural activity but is also a vasodilator [97]. This is likely a survival mechanism, in which adenosine overproduction is linked to a lack of recovery of ATP, possibly from an energy crisis due to a lack of oxygen, and the response would tend to both increase the delivery of oxygen and reduce the demand for oxygen. A second example of opposite effects on neural activity and CBF is nitric oxide, a very potent vasodilator, and yet it is produced by inhibitory interneurons [107]. The presence of these somewhat counterintuitive mechanisms, which drive neural activity (and presumably $CMRO_2$) oppositely to CBF suggests that the coupling of CBF and $CMRO_2$ may be variable.

These considerations of the many potential mechanisms that could be involved in neurovascular coupling highlight the complexity of CBF control. For many of these mechanisms there is a plausible correspondence between flow and energy metabolism demands, in the sense that the aspects of neural activity that are likely to cost the most energy also strongly drive CBF changes. This suggests the possibility that the essentially feedforward mechanisms of CBF control are tuned during development to provide sufficient O_2 to meet the energy demands while maintaining tissue oxygenation, as suggested above. However, the independence of these mechanisms (i.e. the lack of a rapid feedback signal from tissue O_2 to adjust CBF) means that we should treat the $CMRO_2$/CBF coupling ratio λ as a physiological quantity that could vary between brain regions, with different types of stimulus, and with disease. As this parameter strongly affects the BOLD response, the interpretation of the magnitude of the BOLD response in a physiologically meaningful way is difficult.

4. The physical basis of fMRI

This section reviews the physical basis of the BOLD effect, and in particular the quantitative modeling that can be used to understand and interpret the BOLD response. While the modeling reveals the complexity of the BOLD effect, it also provides a theoretical framework for combining different measurements to estimate important physiological quantities. A major ongoing goal of this work is to develop methods for separating the effects of CBF and $CMRO_2$ and thus provide tools to make fMRI into a quantitative probe of brain physiology. These ideas are discussed in more detail in several recent reviews [21, 108–110]. Here the focus is on signal decay, the dominant mechanism exploited in fMRI. However, the magnetic field offsets produced by magnetized blood vessels also interact with a rapidly pulsed method called *steady-state free precession* (SSFP). While the signal

behavior is more complicated than what is described below, this is nevertheless a sensitive and promising approach for detecting and mapping brain activity [111, 112].

4.1. T_2^* decay

After an excitation RF pulse, the measured MR signal is proportional to the transverse magnetization at the time of measurement. The central physics for understanding the BOLD effect is how the decay of that signal is altered by deoxyhemoglobin in the blood vessels. In a typical fMRI experiment images are repeated, each acquired with the same fixed delay TE after excitation. If the deoxyhemoglobin content decreases, the signal decays less and is thus slightly stronger during neural activation. In the following, we consider current physical models for decay of the transverse magnetization relevant for understanding the BOLD effect, including the effects of fluctuating fields, static field inhomogeneities, and diffusion through inhomogeneous fields.

The decay of the signal, starting at the initial excitation at $t = 0$, can be described by an attenuation factor $A(t)$, with $A(0) = 1$. That is, $A(t)$ represents the fraction of the original signal remaining at time t. Note that the actual signal is $A(t)$ scaled by the proton density and potentially by factors related to the repetition time and T_1—here the goal is just to model the decay of the transverse magnetization. In classic NMR theory the attenuation factor for an SE experiment is a simple exponential decay, $A(t) = \exp[-R_2 t]$, where $R_2 = 1/T_2$ is the transverse relaxation rate. For the GE experiment typical of fMRI (i.e. measuring an FID), the signal decay is more complicated because of the effects of microscopic magnetic field distortions around blood vessels containing deoxyhemoglobin. This enhanced decay rate is described by a relaxation rate R_2^* ($>R_2$), although in general R_2^* is a function of time. That is, the attenuation factor is still modeled as nominally an exponential $(A(t)=\exp[-R_2^*(t)t])$, but with time-dependent R_2^*, so the shape of $A(t)$ can depart significantly from a simple exponential. The inverse of R_2^* is the decay time T_2^* ($<T_2$).

Now consider an otherwise uniform tissue element containing blood vessels occupying a volume fraction V (typically ~0.05), and focus on the net signal from the extravascular space, S_E. If the signal from a tissue element with no blood vessels ($V = 0$) is defined to be 1 at $t = 0$, the extravascular signal is

$$S_E = (1 - V)A(t). \quad (2)$$

If there is no difference between the magnetic susceptibility of the blood vessels and the surrounding tissue, the relaxation rate is assumed to be simply $R_2^* = R_2$. Now introducing a susceptibility difference between the blood vessels and the surrounding space, the form of R_2^* can be written as

$$R_2^*(t) = R_2 + \Delta R_2(t) + R_2'(t). \quad (3)$$

The significance of this form is that it breaks the added relaxation into two terms: a part that can be refocused with an SE experiment (R_2') and a part that cannot be refocused and so appears as a change in R_2 (ΔR_2).

The physical effect that determines the balance of the two additional terms in equation (3) is the diffusion of water molecules. If the water molecules were static (no diffusion), then the local magnetization would simply precess at the local Larmor frequency, which is proportional to the local field offset. The net signal would then decay more quickly as the magnetization from microscopic domains precessed at slightly different frequencies producing a growing phase dispersion. In this *static dephasing* limit all of the additional relaxation is in R_2' ($\Delta R_2 = 0$), and it depends just on the overall distribution of field offsets around the vessels. After an evolution time *t* the local phase of each microscopic domain is simply proportional to *t*, and an SE will cleanly reverse the local phase accumulation at each point.

With random motions of the water molecules due to diffusion, though, the net phase of each precessing magnetization at time *t* reflects the past history of locations of the water molecules. That is, the local magnetization precesses at different rates as spins undergo a random walk through the distorted magnetic field. In the *fast diffusion* limit (or the *motional narrowing* regime) water molecules diffuse during the decay time over large regions compared with the spatial scale of the magnetic field distortions, so that effectively each spin samples all of the field offsets created by the magnetized vessels. None of this random phase accumulation can be reversed by an SE, and so in this limit all of the effect of the local field inhomogeneities is in the ΔR_2 term. Note though that because each spin is, on average, sampling all of the field offsets, there is less dispersion of the accumulated phase across spins (the origin of the description *motional narrowing*, as the final distribution of phases is narrower than it would be without diffusion). For this reason, a given amount of deoxyhemoglobin in larger blood vessels, where diffusion effects are less important, has a greater effect on R_2^* than when the same amount of deoxyhemoglobin is in capillaries, where diffusion considerably diminishes the effect on R_2^*. Much of the effort in modeling the BOLD effect has been devoted to exploring the middle ground between the two limits of static dephasing and fast diffusion.

4.2. Magnetic field distortions around a magnetized cylinder

The field distortion around magnetized blood vessels is usually modeled in terms of the magnetic field offset at a point near an infinitely long cylinder with radius *R* oriented at an angle α to the main magnetic field $\boldsymbol{B}_0$ and with magnetic susceptibility difference $\Delta\chi$ between the cylinder and the surrounding space:

$$\Delta B(r,\theta) = \Delta\chi B_0 \left(\frac{R}{r}\right)^2 \sin^2\alpha\cos 2\theta. \quad (4)$$

Note that the field offset in equation (4) is the *z*-component of the added field due to the magnetized cylinder, because this is the component that adds to the much larger main magnetic field B_0 and so alters the precession frequency. Here (r, θ) is a coordinate system oriented to the cylinder, so that *r* is the perpendicular distance from the point to the axis of the cylinder, and θ is the angle of the point with $\theta = 0$ taken as the angle of projection of the main magnetic field $\boldsymbol{B}_0$ into the cylinder space (figure 9). The field distortion is maximized when the cylinder is perpendicular to $\boldsymbol{B}_0$ ($\alpha = \pi/2$), and there is no distortion when it is

parallel to $\boldsymbol{B}_0$ ($\alpha = 0$). There are also two lines at angles $\theta = +\pi/4$ and $-\pi/4$ along which the field is not distorted, giving the overall field distribution the characteristic dipole shape shown in figure 9. Note also that the field falls off with the square of distance scaled by the radius of the cylinder, but that the field offset at the surface of the cylinder ($r = R$) is the same for vessels of all sizes, depending just on geometrical orientation and the susceptibility difference $\Delta\chi$.

The magnetic susceptibility difference is usually taken as $\Delta\chi = (1 - Y)\mathrm{Hct}\Delta\chi_0$, where Y is the fractional O_2 saturation of hemoglobin, Hct is the hematocrit (the volume fraction of blood occupied by red blood cells), and $\Delta\chi_0$ is the susceptibility difference that would result if all of the hemoglobin was deoxygenated ($Y = 0$) and the hematocrit is one. Recent models [113, 114], though, have modified this in keeping with the experiments of Spees *et al* [115] who found that the point of equal susceptibility difference between plasma and red blood cells is not 100% saturation but rather ~95% saturation due to the different protein contents of plasma and red cells. Assuming the susceptibility of the extravascular space is the same as that of the plasma, this effect should be included. For the discussion here we will ignore this correction, and assume the simpler relation between Y and $\Delta\chi$. The offset of the angular frequency of precession is $\gamma\Delta B$, and this distribution determines the accumulated phase that leads to $A(t)$. Taken together, the added angular frequency (rad s^{-1}) distribution around the vessel is

$$\Delta\omega(r,\theta)=\delta\omega_0\left(\frac{R}{r}\right)^2\sin^2\alpha\cos 2\theta \quad (5)$$

with

$$\delta\omega_0=2\pi\gamma B_0\Delta\chi_0\mathrm{Hct}(1-Y). \quad (6)$$

Note that $\delta\omega_0$ is the maximum angular frequency offset at the surface of a cylinder oriented perpendicular to the magnetic field. For a susceptibility difference of $\Delta\chi_0 = 0.264 \times 10^{-6}$ [115], a typical hematocrit of 0.44 (normal range is about 0.42–0.52 for adult males and 0.35–0.47 for adult females), and a baseline venous saturation of $Y = 0.6$, a typical value of $\delta\omega_0$ is ~230 rad s^{-1} for $B_0 = 3$ T. In their seminal paper modeling the BOLD effect, Ogawa and co-workers expressed $\delta\omega_0$ as an equivalent frequency ν (in Hz, rather than rad s^{-1}, so $\nu = \delta\omega_0/2\pi$), which would be ~37 Hz for this example. This estimate for a field strength of 3 T is consistent with assumptions in two recent models [113, 114], but lower values have been assumed in other models based on either lower assumed values of Hct, Y or $\Delta\chi_0$ [116, 117]. Because of physiological variability in the hematocrit value in the normal healthy population, this basic parameter could vary in practice by up to 30%, so there is no definitive value that can be assumed to be accurate for all subjects.

The parameter $\delta\omega_0$ (or ν) is the fundamental scaling parameter for magnetic field offsets around a magnetized cylinder, and the geometry of those offsets is governed by equation (5). For modeling blood vessels in the brain, the usual assumption is that the vessels are randomly oriented. This is probably a reasonable assumption for the smallest vessels, but larger vessels may be more oriented. In the cortex, blood vessels spread over the surface of

the brain with arterioles diving perpendicular to the surface into the cortex [118]. Detailed representations of the human brain vascular network are now becoming available as models or as direct vascular images [119–123], and it will be important to do similar calculations of T_2^* effects on these realistic networks to test whether the idealized assumptions of the numerical models (randomly oriented, infinitely long cylinders) introduce significant systematic errors.

4.3. Static dephasing

Yablonskiy and Haacke [124] analyzed the extravascular signal in the static dephasing regime (no diffusion) for randomly oriented magnetized cylinders occupying a volume fraction *V*. In this regime all of the added relaxation from equation (3) is due to R_2'. It turns out to be useful to express the decay relative to an external medium that contains no vessels ($V = 0$). Then the attenuation curve, rather than starting at a value of one, starts at a value of $1 - V$ (as in equation (2), because the vessels themselves are assumed to generate no signal —intravascular signal changes will be considered later). The signal then gradually decays, initially equivalent to an R_2' value that increases with time (figure 10). As time increases R_2' approaches a value of

$$R_2' = \frac{2}{3} V \delta\omega_0. \quad (7)$$

This form nicely explained the earlier Monte Carlo simulation results of Ogawa *et al* [116]. The numerical factor comes from the random orientation of vessels, effectively creating an average scaling frequency of $\delta\omega_{av} = 2/3\delta\omega_0$. This monoexponential behavior develops for time $t > 1/\delta\omega_{av}$. Importantly, the monoexponential portion extrapolates back to a value of 1 at $t = 0$, so that the ratio of the actual signal at $t = 0$ to the back-extrapolated value from the monoexponential regime depends just on the volume fraction.

We can understand this behavior intuitively by considering two aspects of the field distortions described by equation (5). First, the field offsets (measured as equivalent angular frequency offsets) have a finite range $\delta\omega_0$; that is, the largest offsets are at the surface of the cylinder, and depend only on the susceptibility difference and not the cylinder radius. Second, if we consider a cylinder with radius *R*, the field distribution for $r > R$ is identical to the field offsets produced by a thinner cylinder with a larger susceptibility difference, provided the product $R^2\delta\omega_0$ remains constant. For the assumed distribution of infinite cylinders, the blood volume fraction *V* translates to R^2, so this basic combination of physical variables, $R^2\delta\omega_0$, is at the heart of equation (7). We can think of the extravascular signal as arising from a series of thin shells around the magnetized cylinder. As time increases, the phase variation of the signals within a shell grows, and we can loosely think of the net signal from a shell as being destroyed once the phase variation reaches a sufficiently high level. Note, though, that once the signal from a shell near the surface of the cylinder has been destroyed, information on the separate values of *R* and $\delta\omega_0$ is lost, because the field distortions in the remaining space generating a signal are the same for any set of cylinders for which $R^2\delta\omega_0$ is constant.

Note that this is not true, though, for the earliest part of the signal decay when the signal from the shell closest to the cylinder has not been destroyed. Initially, the extravascular signal is 1 − υ, and as time increases to the point that the signals from the shells closest to the cylinder are destroyed, the subsequent signal decay is the same as it would be for a set of very thin cylinders with a large susceptibility difference such that $V\delta\omega_0$ remains the same. For such cylinders, the signal from the inner shells is destroyed almost immediately, so the projection back to $t = 0$ is to a signal of 1 (i.e. it approaches the condition of $V = 0$). For this reason, the projection of the decay curve back to zero provides information on the blood volume fraction V independent of the R_2^* value observed at later times, which depends only on the product of V and $\delta\omega_0$.

Potentially, this provides a way of disentangling the effects of blood volume and blood susceptibility. This is critical, because $\delta\omega_0$ is the parameter that directly relates to blood oxygenation and to the OEF, and a robust measurement of OEF provides information on oxygen metabolism. Put another way, blood volume is a confounding effect such that R_2^* is not a clean reflection of OEF alone, and so the ability to separately account for blood volume effects is critical for quantifying $CMRO_2$ from BOLD measurements. This remains a challenge, although Yablonskiy *et al* [110] have developed an approach called qBOLD based on this key insight related to extravascular static dephasing and the shape of the decay curve (discussed further below).

4.4. Diffusion effects

Modeling the effects of diffusion is considerably more complicated, and nearly all studies are limited by the need to assume relatively simple geometries. Several analytical treatments have been proposed [125–128], and a number of studies have used Monte Carlo simulations tracking thousands of random walks through a field of randomly oriented vessels to model diffusion [114, 116, 129–131]. In thinking about diffusion, the critical characteristic time is the time required for diffusive motions to be comparable to the spatial scale of the field distortions, $\tau_D \sim R^2/D$, where R is the radius of the blood vessel creating the field offsets and D is the diffusion coefficient. If the evolution time of the signal is much longer than τ_D, then each diffusing spin will tend to sample all of the field offsets produced by the magnetized vessel. At this extreme (the *fast diffusion* limit) we can estimate the expected behavior with a random walk argument. Let $\delta\omega^2$ be the mean squared field offset in the space around the magnetized cylinder within a radial distance r_{max}. This maximum radial distance is determined by the blood volume fraction V (essentially we are averaging out to the point of overlap of field offsets due to neighboring blood vessels), with $V = (R/r_{max})^2$. For parallel cylinders with low volume fractions, $\delta\omega^2 \sim V\delta\omega_0^2/2$ [127]. For the fast diffusion limit, we consider each spin to be randomly sampling field offsets from a Gaussian distribution with variance $\delta\omega^2$, and precessing for a time τ_D in each field, so that in each step it acquires a phase (squared) of $\delta\phi^2 = \tau_D^2\delta\omega^2$. After evolving for a time t ($N = t/\tau_D$ steps), the variance of the accumulated phase is $\sigma_\phi^2 = N\delta\phi^2 = t\tau_D\delta\omega^2$. When this Gaussian distribution of phase is multiplied by cos(φ) and integrated to calculate the decay curve, the result is a monoexponential decay for fast diffusion with ΔR_2 of the form

$$\Delta R_2 = \kappa \tau_D V \delta\omega_0^2 \quad (8)$$

with κ a dimensionless numerical constant (for this simplified estimate, κ = 0.25).

Equations (7) and (8) essentially bracket the range of effects on the signal decay curve due to blood vessels with blood volume fraction *V* and a susceptibility difference characterized by $\delta\omega_0$. With no diffusion (static dephasing) the added decay is described entirely by R_2' (equation (7)), with a linear dependence on $\delta\omega_0$. In the fast diffusion limit, the decay is described by ΔR_2 (equation (8)), with a quadratic dependence on $\delta\omega_0$.

In practice, the decay of the signal in the BOLD effect falls between these two limits for the smallest blood vessels (e.g. for a capillary with a radius of 2.5 μm and $D \sim 1$ μm^2 ms^{-1}, the characteristic time is $\tau_D \sim 6$ ms), and so a number of investigators have approached this problem with Monte Carlo simulations. Figure 10(*b*) shows Monte Carlo simulations from Dickson *et al* [131] for the change in relaxation rate for a GE experiment with TE = 30 ms and SE experiment with TE = 60 ms as a function of blood vessel size. (If these simulations had been performed for the same TE, the SE curve would correspond to ΔR_2 in equation (3), and the difference between the SE and GE curves would correspond to R_2'.) Note that from these calculations we would expect vessels with $R > 10$ μm to be solidly in the static dephasing regime, and equation (7) is typically used to model the effects of these larger vessels. However, the smallest capillaries with $R \sim 2.5$ μm are part-way toward the fast diffusion limit, with some contribution to R_2^* from both ΔR_2 and R_2'. For an assumed diffusion constant of $D = 1$ μm ms^{-1}, Ogawa *et al* [116] found that the effects of the largest vessels were accurately described by equation (7), while the effect of the smallest vessels ($R = 2.5$ μm) could be described by equation (8) with the product $\kappa\tau_D = 0.0010$ s.

A key result of the arguments above is that diffusion always tends to reduce the BOLD effect compared with what it would be in the absence of diffusion. Physically, this is because any distribution of field offsets produces the widest distribution of phases when there is no diffusion, because diffusion effectively brings in averaging over the field distribution (*motional narrowing* of the phase distribution) as described earlier. The magnitude of this effect can be estimated from equations (7) and (8) by considering two blood vessel distributions, one of larger vessels ($R > 10$ μm) and one of capillaries ($R = 2.5$ μm), each with the same blood volume fraction and each with the same deoxyhemoglobin content (equal $\delta\omega_0$). Curves showing the signal loss due to each population calculated for $\delta\omega_0 = 200$ rad s^{-1} at a field of 3 T are plotted in figure 10(*c*), showing an approximately three-fold difference for an O_2 extraction fraction of 40%. In addition, the hemoglobin oxygen saturation is expected to be higher on average for capillaries compared with veins, as it varies down the length of the capillary from a value near the arterial saturation to the venous value. Taken together, these arguments suggest that the standard BOLD effect measured in GE imaging is dominated by the veins, with a much smaller contribution from the capillaries.

However, the physical effects of diffusion also suggest another possible application: using an SE experiment instead of a GE experiment could isolate signal changes related to the

smallest vessels. That is, the SE experiment is only sensitive to the ΔR_2 effects (the R_2' effects are refocused), and these effects only occur for the smallest vessels (figure 10(*b*)). The significance of this idea is related to an issue noted early in the development of fMRI and referred to as the *brain versus vein* problem [132, 133]. The largest change in deoxyhemoglobin is in the veins, but a draining vein may show a significant BOLD effect that is displaced from the site of the neural activity change. The potential of SE-BOLD imaging is that such draining veins would contribute less to the net signal than the effects due to the capillaries, and as a result the signal would be better localized to the true site of the neural activity change [134]. However, because the SE-BOLD effect is so much weaker than the GE-BOLD effect, this approach is not practical except at high main magnetic fields where the SNR is significantly improved [31, 135].

In addition, the argument for the better selectivity of the SE-BOLD experiment is based on considerations of the extravascular signal change. In practice, at moderate field strengths (1.5–3 T), the intravascular signal change is significant, even for an SE experiment [136, 137]. At high fields the intravascular contribution is reduced [138], so again the idea of using SE-BOLD for better localization is only practical for high fields. Interestingly, a recent modeling study suggests that the SE-BOLD sensitivity to the microvasculature may peak at very high fields, and that the SE sensitivity to the vascular signal of capillaries and arterioles persists even at high field strengths [114].

The different sensitivities of GE and SE imaging to vessel size also suggests the possibility of vessel size imaging, by looking at the ratio of ΔR_2 and ΔR_2^* [139–141]. This ratio steadily diminishes as the vessel size increases, and can provide a way to investigate angiogenesis, among other applications.

4.5. Intravascular BOLD effect

The arguments above focused on the signal changes in the extravascular space, and did not consider changes in the signal generated by blood itself (as noted in the last section). The intravascular compartment is a small fraction of the total tissue volume (only about 5%), and so at first glance one might suppose that the intravascular spins would contribute a comparably small amount to the net BOLD signal change. However, at 3 T and below, the vascular contribution is comparable to the extravascular contribution [142–144]. Water molecules in blood are closer to the source of the magnetic susceptibility change (deoxyhemoglobin), significantly increasing the effect, but this is somewhat moderated by strong diffusion effects as well [12]. The net result is that the intrinsic signal change in the blood is more than an order of magnitude larger than the extravascular signal change [142]. Figure 10(*d*) shows curves of the R_2^* of blood based on empirical data measured at 3 T for various hematocrit values [145] and fit to a simple quadratic function that includes the hematocrit as a potential variable [113]. Comparing extravascular and intravascular signal changes from the curves in figures 10(*c*) and (*d*), an activation that reduced the OEF from 40% to 30% would create an extravascular signal change of ~2% and an intravascular signal change of ~50%. Even with a blood volume fraction of only 0.02, the intravascular signal change still accounts for about one-third of the total signal change due to its much wider dynamic range.

4.6. Modeling the BOLD effect

The previous sections have described individual components of the BOLD signal, including static dephasing effects, the role of diffusion and the contribution of intravascular changes. In practice, the measured signal change also includes a volume exchange effect, as CBV increases and displaces extravascular water generating a different signal. While this effect is not 'blood oxygenation dependent', it nevertheless contributes to the measured signal and should be included in the modeling [146]. A full quantitative model of the BOLD effect is important for understanding the basic mechanisms, for optimizing the image acquisition technique to maximize sensitivity, and for calibrating the BOLD signal to measure local $CMRO_2$. The latter calibrated BOLD application (described in more detail in the next section) is particularly important, because it provides a way to measure the $CMRO_2$ response to a stimulus [108, 109]. The basic idea is that the BOLD response depends on both the CBF and $CMRO_2$ responses. If the BOLD response is measured in conjunction with an ASL experiment to measure the CBF response independently, one can in principle isolate the $CMRO_2$ response. The key to making this approach work is the availability of an accurate model relating the BOLD signal change to the underlying changes in CBF and $CMRO_2$. This section reviews modeling efforts in this direction, focusing on the GE-BOLD signal at 3 T as the most commonly used technique. Uludag *et al* extended these modeling ideas in a comprehensive way to include SE-BOLD and a wide range of field strengths [114].

Historically, the most influential model of the BOLD effect was introduced by Davis *et al* in their seminal paper describing the calibrated BOLD approach [130]. In the context of the terminology used in the current review (equation (3)), they modeled the combined term $\Delta R_2 + R_2'$, which can be abbreviated as R_2^*(dHb), the additional contribution to the relaxation rate due to the presence of deoxyhemoglobin. They assumed the form

$$R_2^*(\mathrm{dHb}) = kV[\mathrm{dHb}]^{\beta} \quad (9)$$

where V is the blood volume fraction, [dHb] is the local concentration of deoxyhemoglobin in blood, and k is a constant that depends on field strength. They suggested a value of $\beta = 1.5$ as a compromise between the effects of large and small vessels (as in equations (7) and (8), where the exponent on $\delta\omega_0$ varies from 1 to 2). The fractional BOLD signal change between a baseline state (denoted by subscript '0') and an activation state (no subscript) is due to the change in R_2^*(dHb) between those two states. In these calculations of a fractional signal change ($\Delta S/S_0$), the effects of R_2 cancel out, so this term is dropped from the following equations. The measured GE signal is

$$S \propto \mathrm{e}^{-R_2^*(\mathrm{dHb})\mathrm{TE}} \cong 1 - R_2^*(\mathrm{dHb})\mathrm{TE} \quad (10)$$

where TE is the echo time of the experiment, and the approximation is based on the assumption that the exponent is small compared with one. The fractional BOLD signal change between a baseline state and an activated state is then

$$\frac{\Delta S}{S} \cong kTE(V_0[\mathrm{dHb}]_0^\beta - V[\mathrm{dHb}]^\beta) = kTEV_0[\mathrm{dHb}]_0^\beta(1 - \upsilon c^\beta) \quad (11)$$

where lowercase symbols are the values in the activated state normalized to the baseline state values (i.e. $\upsilon = V/V_0$ and c is the ratio of [dHb] in the active state to the baseline state). The deoxyhemoglobin concentration in blood ([dHb]) was modeled as though it was all in the venous vasculature, so that the concentration ratio is just the ratio of OEFs (E/E_0), and equation (1) can be used to express this in terms of the CBF normalized to its baseline value (f) and the $CMRO_2$ normalized to its baseline value (r). Combining the multiplicative constants into a single scaling factor M, the BOLD signal is

$$\frac{\Delta S}{S} = M\left[1 - \upsilon\left(\frac{r}{f}\right)^\beta\right]. \quad (12)$$

Finally, the blood volume change is assumed to be related to the blood flow change with a power law relationship, $\upsilon = f^\alpha$, with $\alpha = 0.38$ based on experiments comparing total CBV with CBF [147], so that the final model is

$$\frac{\Delta S}{S_0} = M(1 - f^{\alpha-\beta}r^\beta). \quad (13)$$

This simple form satisfies the goal of relating the BOLD signal change to the changes in CBF (f) and $CMRO_2$ (r), and has since been widely used. Note, though, that there are some issues with this model. The derivation of this expression leaves out intravascular signal changes and volume exchange effects. In addition, in recent years the question of the right value of α has been raised, as a number of experiments have found that the volume change on the venous side is smaller than the overall blood volume change [148–152]. The right value of β also has been questioned, and as fMRI studies have moved to 3 T fields, an assumed value of $\beta = 1.3$ is becoming more common [153]. A subsequent model from my group (originally in [154] and then modified in [146, 155]) included intravascular signal changes and volume exchange effects. This model did not have an exponent β and involved two constants, rather than the single M in the Davis model.

Recently, more detailed models have dispensed with the goal of a simple final form and included all of the effects currently thought to be contributing to the BOLD response [113, 114]. Griffeth and co-workers in my group used such a detailed model to test the accuracy of the Davis model and found that, despite the limitations of the original derivation, the form of the Davis model is reasonably accurate [113]. This is because many effects that simply scale the overall BOLD signal are captured by the scaling constant M. Further analysis with the detailed model currently underway suggests that an alternative heuristic model is approximately as accurate as the Davis model:

$$\frac{\Delta S}{S_0} = A(1 - \alpha_V - \lambda)\left(1 - \frac{1}{f}\right) \quad (14)$$

where α_V is specifically the exponent for the volume change of the venous compartment (e.g. 0.2 based on [152]), and λ is the ratio of the fractional $CMRO_2$ change to the fractional CBF change (e.g. for a 20% CBF change accompanied by a 10% $CMRO_2$ change, $\lambda = 0.5$). The heuristic advantage of equation (14) is that it clearly shows the range of effects that modulate the BOLD signal. A number of factors primarily just scale the BOLD response, and so are subsumed in a scaling factor *A*. A venous volume increase with activation tends to reduce the BOLD effect by increasing the total deoxyhemoglobin, reflected by the α_V term. The change in blood oxygenation strongly depends on the relative changes in CBF and $CMRO_2$, reflected in the λ term. Finally, there is an intrinsically nonlinear dependence on the CBF change, reflected in the last term. This nonlinearity is due to the ceiling effect on the BOLD signal: the maximum that a large CBF increase can do is to reduce deoxyhemoglobin to near zero values, and this would correspond to a finite signal change [156].

5. Functional MRI as a quantitative probe of brain physiology

5.1. The calibrated BOLD method

A key feature of the basic BOLD signal models is that they include an overall scaling parameter (*M* in equation (13) and *A* in equation (14), and it will be discussed here as the more familiar *M*). This is critical because *M* captures much of the physiological variability of the baseline state. That is, *M* depends on the amount of deoxyhemoglobin in the baseline state, which is affected by hematocrit, venous blood volume and the baseline OEF. As a result, *M* should be assumed to vary across individuals, across brain regions, and even within the same individual brain region if the baseline state changes. The downside of this is that *M* must be measured in each subject, but the upside is that once it is measured it accounts for much of the variability across subjects. In addition, knowing *M* makes it possible to interpret changes in the BOLD response in terms of changes in $CMRO_2$ when CBF is measured concurrently with an ASL technique. For example, from equation (13), if *f* is known from the ASL measurement and *M* is known in some way, then the measured BOLD signal change can be used to calculate *r*, the ratio of the $CMRO_2$ value in the active state relative to the baseline state. The key requirement then is a method for determining *M*, called the BOLD calibration.

Davis *et al* [130] proposed the calibrated BOLD approach based on measuring the CBF and BOLD responses to inhaling a gas containing 5% CO_2 (hypercapnia). Based on a number of studies, mild hypercapnia was thought to produce a large increase in CBF with little or no change in $CMRO_2$, and so $r = 1$ is assumed. Then from equation (13), with *f* known from the ASL experiment, the value of *M* can be estimated. This has proven to be a powerful tool for exploring $CMRO_2$ changes, particularly for exploring how the $CMRO_2$/CBF coupling ratio λ varies under different conditions (for recent reviews see [108, 109, 157]).

However, in recent years several studies have questioned whether inhaled CO_2 really has no effect on $CMRO_2$ [158–161]. At high levels, CO_2 acts as an anesthetic, so we would expect $CMRO_2$ to decrease for high levels of inhaled CO_2. The essential question, which is still debated [162], is whether modest levels (5% CO_2) also have an effect on $CMRO_2$. In part because of these concerns, there has been an effort to develop alternative methods for

calibration. One suggestion is to use hyperoxia instead (i.e. breathing a gas mixture with 50% or more O_2) [153, 163]. In normal arterial blood the hemoglobin saturation is near 100%, so hyperoxia primarily increases the O_2 concentration as dissolved gas in the plasma. As the hyperoxic blood reaches the capillaries, the excess dissolved O_2 diffuses into the tissue and offsets some of the O_2 that previously came from the hemoglobin-bound pool. As a result, the venous deoxyhemoglobin concentration decreases slightly, creating a BOLD signal increase. With appropriate assumptions about the physiological variables, an estimate of *M* can be derived. Recently, Blockley and co-workers from my group analyzed both the hyperoxia and hypercapnia methods with our detailed model of the BOLD response, and found that the hyperoxia method was sensitive to the assumed value of baseline OEF [164]. Given that variability of baseline OEF is one of the key parameters that *M* is designed to capture, this suggested that hyperoxia may not be robust for determining *M*. In addition, a recent study found evidence that hyperoxia also may alter $CMRO_2$ [165].

Blockley and co-workers also proposed a different way of estimating *M* without using inhaled gases, based on a measurement of R_2' in the baseline state. That is, R_2' can be measured with appropriate comparisons of GE and SE signals. The assumption is then that R_2' is due just to local deoxyhemoglobin. While this approach does not capture the intrinsic ΔR_2 contributions to the net change in R_2^* with activation, those changes are likely to be a minor fraction of the signal change due to R_2'. Put another way, R_2' in the baseline state is primarily determined by total deoxyhemoglobin, as is *M*. By modeling the experiment with the detailed BOLD model one can estimate a scaling factor needed to derive the full value of *M*. However, this approach is still speculative. A critical technical problem is that in practice R_2' is contaminated by field inhomogeneities that produce additional signal loss, so these field variations must be accurately measured and corrected to derive an estimate of R_2' due to deoxyhemoglobin [166, 167]. Nevertheless, the promise of this approach is a much simpler calibration procedure that does not require inhalation of special gases.

In summary, the calibrated BOLD approach is a powerful way to measure fractional changes in $CMRO_2$, which are very hard to measure with any technique [19]. There are still concerns about whether hypercapnia is isometabolic as used in the calibration experiment, and whether the baseline OEF is sufficiently stable across the population (particularly in disease states) so that the hyperoxia method can be robust with an assumed value of OEF. Simpler methods of calibration based on R_2' in the baseline state look promising, as they do not require inhalation of special gas mixtures, but it remains to be demonstrated that the confounding effects of field inhomogeneities on R_2' can be controlled sufficiently. Promising approaches are to use other methods for determining some of the model parameters, such as the blood volume effects, in combination with BOLD-fMRI and ASL measurements [144, 168–170].

5.2. Interpreting the BOLD response

The complexity of the BOLD effect described in the previous sections leads to a fundamental problem for interpreting the magnitude of the BOLD response in a quantitative

way in terms of the underlying physiology. Two recent studies illustrate this problem. In the first example, Fleisher *et al* [171] compared the BOLD response in the hippocampus during a memory task between two groups, one consisting of subjects with specific risk factors for the development of Alzheimer's disease, and the other an age-matched control group without those risk factors. They found a weaker BOLD response in the higher risk group, suggesting an early effect of the disease. The interesting question, though, is how should this result be interpreted in terms of the underlying physiological effects? It is tempting to interpret this as a sign that the neural processing associated with the memory task was altered. However, the model in equation (14) shows that for the same change in oxygen metabolism, the BOLD response could differ if the $CMRO_2$/CBF coupling ratio λ changes, or if the baseline state changes affecting A. In particular, the scaling parameter A essentially reflects the deoxyhemoglobin content in the baseline state, which can be altered by changes in baseline O_2 extraction fraction or venous blood volume. The latter effect is particularly a concern in looking at patient populations, where chronic changes in the baseline state due to the disease or the effects of medications could alter the BOLD response to a standard task, even though the CBF and $CMRO_2$ responses to the task are unchanged. In the Fleisher study they also used ASL to measure CBF in both the baseline state and the active state, with the interesting finding that CBF in the active state was similar in the two groups, but the baseline state was different. This suggests that the change in the acute BOLD response to the task may be partly due to a chronic difference in the baseline state.

A second example used dual measurement of ASL and BOLD responses in a calibrated BOLD method to untangle the effects of caffeine on both baseline state changes due to the drug and alterations in the way the brain responds to a visual stimulus with the drug on board [21]. Interestingly, the BOLD response to the visual stimulus was nearly identical pre- and post-caffeine. Nevertheless, the underlying physiological changes measured with the calibrated BOLD method were substantial: baseline CBF decreased while baseline $CMRO_2$ increased, and in response to the visual stimulus the $CMRO_2$ response increased by more than 60%. These findings are consistent with the known action of caffeine in blocking adenosine receptors [98], and with other fMRI studies with caffeine [172]. Because adenosine both increases blood flow and acts as an inhibitor of neural activity, when adenosine receptors are blocked we expect CBF to go down, but because inhibition of neural activity is partially lifted we expect $CMRO_2$ to go up. The fact that the BOLD signal failed to reflect these underlying changes is due to the conflict between two effects: the baseline state changes should increase the baseline OEF, which would tend to increase the BOLD response; and the larger ratio of the change in $CMRO_2$ to CBF in response to the visual stimulus, which tends to decrease the BOLD response. In terms of equation (14) this amounts to an increase in both A and λ, which have opposing effects on the overall scaling of the BOLD signal, and in these data the two effects balanced out to produce no change in the BOLD response. In short, while a physiological interpretation of the BOLD signal alone is not possible, the addition of other measurements makes a much more quantitative probe of brain physiology possible.

5.3. Dynamics of the BOLD response

The time scale of the BOLD response is much slower than the time scale of neural activity because the CBF response develops and recovers relatively slowly. Even a brief sub-second neural stimulus produces a BOLD response that is delayed by a few seconds and may take about 6 s to evolve [173]. This time scale for the BOLD response may vary across the brain and across subjects, but a more interesting finding is that it can also vary in the same subject depending on the physiological baseline state. In particular, altering baseline CBF slows or speeds up the BOLD response in a counterintuitive way: increased baseline CBF slows the BOLD response, while decreased baseline CBF leads to a faster BOLD response. For example, Cohen *et al* found this pattern when baseline CBF was increased by inhaling a gas mixture with increased CO_2 content (hypercapnia) or reduced by voluntary hyperventilation (hyocapnia), exploiting the intrinsic sensitivity of CBF to the CO_2 content of blood [174]. Other studies with caffeine, which lowers baseline CBF, found faster BOLD response dynamics [175, 176], consistent with this basic trend.

A possible explanation for this phenomenon was proposed by Behzadi and Liu [177] based on a biomechanical model in which the net compliance of the artery is due to two factors: the tension of the smooth muscle, and an elastic component that increases as the vessel is stretched. The essential physical idea of the model is that when the vessel is constricted, so that the elastic components are not stretched, the compliance is dominated by the smooth muscle and responds more quickly to changes in smooth muscle tension. In contrast, when the artery is dilated the elastic components make a significant contribution to the overall compliance of the vessel, and the same relaxation of the smooth muscle would not produce as large a change of the overall compliance.

The dynamics of the BOLD response also exhibits interesting transient features, in particular a brief *initial dip* before the primary positive BOLD response [178] and a much longer *post-stimulus undershoot* after the end of the stimulus [179, 180]. While the initial dip is not always observed, the post-stimulus undershoot is common. Interestingly, there is still no consensus on the physiological origin or significance of these features. While it is possible that these transients reflect underlying transients of the neural response, it is also easy to imagine that they arise from different time constants for the physiological variables that contribute to the BOLD signal: CBF, $CMRO_2$ and venous CBV. For example, if the $CMRO_2$ increases before the CBF begins to change, the BOLD response could show an initial dip due to the increase in deoxyhemoglobin. A post-stimulus undershoot could occur if the flow transiently drops below the baseline level, or if the flow returns quickly to the baseline but the venous blood volume or oxygen metabolism returns more slowly. Because of this dependence of the BOLD signal on multiple physiological changes, it is not possible to identify the sources of these transients from BOLD measurements alone.

The initial dip was first detected in animal experiments with optical methods sensitive to the different reflectance properties of hemoglobin and deoxyhemoglobin [181, 182] and later found in human fMRI studies [183]. The basic finding was an early increase in deoxyhemoglobin, prior to the much larger decrease in deoxyhemoglobin that produces the BOLD effect. The dominant view since the early days has been that the initial dip represents

a quick increase in $CMRO_2$ before the CBF can increase, so that initially the OEF increases. However, the interpretation of the initial dip has been controversial because of potential confounding factors in the optical measurements and because of the possibility of CBV effects in the BOLD signal. Optical studies in animal models have found evidence for both an early change in oxygenation before a blood volume change [184] and for blood volume changes as the earliest effect [185]. An fMRI study showing that caffeine, by quickening the CBF response, also reduces the initial dip is consistent with the idea of a faster rise in $CMRO_2$ compared with CBF [176]. However, Uludag *et al* [114] recently argued that volume exchange effects with the arterial compartment could lead to an initial dip as well. In short, the physiological origin of the initial dip is still unsettled [186, 187]. A strong motivation for understanding the initial dip has been that if it *is* due to an early $CMRO_2$ increase, then mapping the initial dip may provide a more accurate map of the location of neural activation, based on the idea that the $CMRO_2$ response is better localized than the CBF response that dominates the primary positive BOLD response. However, in practice, the weakness of the initial dip makes it challenging to detect except at higher fields [188], and even harder to map accurately.

A post-stimulus undershoot was present in the seminal data from Kwong *et al* [3] demonstrating the BOLD effect in humans in response to a visual stimulus, and yet its interpretation is still controversial (see figure 11 for a recent example). Several mechanisms could potentially produce this effect, and the current experimental data do not provide a clear interpretation (see recent reviews [179, 180]). While the effect could be due to an undershoot of neural activity, the duration of the BOLD post-stimulus undershoot (often >30 s) makes this unlikely as a complete explanation, although there are intriguing suggestions that the undershoot is modulated by neural activity [189]. An early and still leading proposed explanation is that it is due to a slow recovery of $CMRO_2$ after CBF has returned to the baseline, so that the OEF is increased during this period [190, 191]. However, animal model studies using an agent (MION) that stays in the vasculature and alters the MR signal in a way that reflects CBV found a slow recovery of CBV during the undershoot period [15]. This motivated the development of two similar models, the balloon model [154] and the windkessel delayed compliance model [192], as biomechanical explanations in which the venous CBV recovers slowly, even though CBF and $CMRO_2$ have returned to the baseline so that OEF is back to the baseline as well.

Note that these two potential explanations reflect the fundamental ambiguity of BOLD measurements: both venous CBV and OEF affect the BOLD signal. The evidence favoring the slow $CMRO_2$ recovery is primarily the absence of evidence for a slow CBV recovery in other studies using different methods sensitive to CBV. Animal models initially found little evidence for venous volume changes with short and moderate duration stimuli [149], although more recent studies found evidence for slowly increasing venous blood volume with longer stimuli [193, 194]. In human studies using different techniques sensitive to blood volume most have found no evidence to support a slow CBV recovery in the BOLD undershoot period [195–199], although one study using a technique sensitive to venous blood volume did find evidence of a slow recovery [200]. Two recent studies used a breath-hold paradigm based on the idea that this would change CBF but not $CMRO_2$, so there

would be CBV changes but not $CMRO_2$ changes. They found no BOLD post-stimulus undershoot, consistent with slow $CMRO_2$ recovery as the explanation when the undershoot is seen [201, 202]. However, a recent animal study found that increased intracranial pressure reduces the undershoot, consistent with a biomechanical explanation [203]. A third possibility is that instead of being due to a slow recovery of CBV or $CMRO_2$, the undershoot is associated with a slight undershoot of CBF [200]. Although ASL studies sometimes show a small undershoot in the CBF response, they do not reach statistical significance. However, this may simply be a problem with the sensitivity of the ASL measurement. Theoretical calculations [180] suggest that the level of CBF undershoot required to explain the BOLD undershoot is likely to be in the noise of most ASL measurements. Clearly further work is needed to understand the post-stimulus undershoot, and it may be that several of these possible explanations contribute.

5.4. Measuring baseline oxygen metabolism

Currently, we know much more about the behavior of CBF than $CMRO_2$, because it is much harder to measure oxygen metabolism [19]. The calibrated BOLD approach described in the previous section makes it possible to measure the fractional $CMRO_2$ change between a baseline state and an activated state, but this still does not give an absolute measure of $CMRO_2$ in either state. In disease the key physiological change may be a chronic shift in $CMRO_2$, rather than a change in how $CMRO_2$ responds to an acute stimulus [62]. In contrast to the case for $CMRO_2$, for studying CBF the ASL method does make it possible to measure absolute CBF in any physiological state. The development of a comparable robust methodology for measuring absolute $CMRO_2$ in the human brain noninvasively would be an important advance for physiological studies of brain function. The basic question is how can we estimate the local OEF? If OEF is known, $CMRO_2$ can be calculated with an additional ASL measurement of CBF and equation (1).

One approach to measuring local OEF centers on measuring the venous oxygenation in specific vessels, based on the effect of deoxyhemoglobin on either the T_2 of blood or the magnetic susceptibility of blood. In the relaxometry approach, the signal of blood in a venous vessel is isolated with a variation of an ASL method, and a multi-echo acquisition then provides a measure of the R_2 of the blood [204–209]. With a calibration curve appropriate for the experiment, the measured R_2 is converted to venous O_2 saturation [210]. The deoxyhemoglobin in venous blood also alters the magnetic susceptibility of the blood, which alters the magnetic field in and around the vessel. This is essentially the macroscopic counterpart to the BOLD effect, in that the BOLD effect arises from field distortions that are too small to measure directly, but their effect is manifested in the signal decay curve. For a large enough venous vessel the susceptibility-induced phase of the blood can be resolved in the MR image and measured directly and then related to blood oxygenation [160, 211–213]. While these methods have primarily been used for essentially whole brain measurements, this work extends these methods to multiple vessels for more localized measurements [214]. In principle, this approach is only limited by the resolution of the images, although the question of determining the drainage basin for individual vessels for accurate localization of $CMRO_2$ remains an issue.

Two alternative approaches for localized $CMRO_2$ measurements are directly related to the BOLD modeling considerations in the earlier sections. In general, the basic problem in estimating baseline OEF from BOLD-related measurements is that the relaxation effects primarily depend on the *total* deoxyhemoglobin within a voxel, and that depends on both baseline OEF and venous blood volume. That is, if we could isolate the parameter $\delta\omega_0$ it would provide an estimate of the baseline OEF, but the basic effect is governed by equations (7) and (8), which depend on venous CBV (V) as well. For these BOLD-based measurements, the challenge then becomes the question of how to estimate the local venous CBV.

The first approach designed to solve this problem is based on a detailed analysis of the full decay curve of the signal, including the initial rounded portion before monoexponential decay begins and the decay around an SE (figure 10(*a*)) [110, 167, 215–217]. As described in section 4.3, for the extravascular signal the initial value is reduced—in proportion to local CBV—from the back-extrapolated value from the monoexponential decay portion of the curve. In short, if the signal decay was purely due to static dephasing in the extravascular space, measuring the full decay curve would provide a way to separately measure V and R_2', and thus isolate $\delta\omega_0$ and estimate OEF. Following the work of Yabonskiy and co-workers, this general approach is usually described under the label qBOLD. Several authors have extended the qBOLD method with sophisticated modeling approaches to dealing with the other factors affecting the signal decay curve (including intravascular effects and diffusion effects) [110, 117, 167, 218]. The full model for the decay curve has become somewhat complex, with a number of parameters describing various physiological effects. The central question with any such model is whether the value of one particular parameter of interest (baseline OEF in this case) can be robustly estimated in the face of physiological variability across the population of subjects. Experimental studies in simple model systems that try to mimic characteristics of a vascular bed found that robust separation of the relevant variables was difficult [219, 220]. A recent Bayesian analysis suggests that very high SNR is required to accurately estimate both blood volume and OEF from qBOLD data [221]. An alternative to trying to derive information on blood volume from the decay curve alone is to combine the basic ideas of qBOLD with an additional measurement of blood volume [166, 222].

An alternative approach has evolved from the work of Gauthier and co-workers [223–225] and Bulte *et al* [226] related to the measurement of M in the calibrated BOLD experiment. Essentially, this approach is the positive aspect of the concerns raised in the previous section about the hyperoxia approach for estimating M due to the strong sensitivity to the assumed value of baseline OEF. The basic idea is that the combination of hypercapnia and hyperoxia experiments may be able to provide an estimate of baseline OEF. As originally developed, the motivation is that both the hypercapnia and hyperoxia experiments should yield the same value of M, *if* the correct value of baseline OEF is assumed in calculating M from the hyperoxia experiment (and the hypercapnia derived M is accurate). Then the assumed value of OEF in the analysis of the hyperoxia data is adjusted until the M-values agree. Another recent work from Blockley *et al* [227] suggests another way of looking at this basic approach, based on the idea that the primary physiological sensitivity of the hyperoxia experiment is really to venous blood volume, rather than M itself. Then the value of M

derived from the hypercapnia experiment depends on both baseline venous CBV and baseline OEF, while the hyperoxia experiment primarily provides a measure of venous CBV, and so the two together provide the information needed to estimate baseline OEF. In short, the basis of the method is that BOLD signal changes with hypercapnia and with hyperoxia depend on venous CBV and baseline OEF in different ways, so it is possible to untangle their effects. This approach is promising, although technically involved because of the need to have subjects breathe special gas mixtures.

6. Summary and future directions

Functional MRI has grown from the conjunction of two phenomena: the biophysical effect that deoxyhemoglobin has magnetic properties that affect the MR signal, and the physiological effect that CBF increases much more than oxygen metabolism as neural activity changes, so that the deoxyhemoglobin level changes as neural activity changes. The resulting BOLD effect has been exploited as a sensitive tool for investigating the working human brain, with the capability of distinguishing subtle differences in patterns of brain activity. The power of a technique that is noninvasive and can be repeated many times for longitudinal studies should not be underestimated, because it makes possible extended studies of the healthy human brain that would not be possible with techniques involving radioactive tracers. Importantly, this means that fMRI can be used to study the dynamics of the human brain in a way that other methods cannot. The limits of spatial and temporal resolution continue to be pushed back, as higher magnetic fields improve the SNR and as technological innovations such as parallel imaging make it possible to reconstruct accurate images from less acquired data.

The challenge for fMRI investigators is to fully understand the depth of information that can be derived from fMRI signals. The BOLD effect alone has primarily been used as a mapping tool (‘Where is the activation?’) and as a qualitative comparison tool (‘Is the response to task A larger than the response to task B in a particular brain region?’). These features alone have led to innovative studies of the integrated dynamics of the brain, based just on how the dynamics of the BOLD signal in one area correlate with the dynamics in other areas. Yet there is still a fundamental problem in the quantitative interpretation of the magnitude of the BOLD response. If the BOLD response to a standard task is larger in one group than another, how should this be interpreted? It is tempting to conclude that the neural activity associated with performance of that task is different, but based on our understanding of the physical origins of the BOLD effect this conclusion is not justified. Systematic differences in the baseline state between groups, such as medications, caffeine intake, or even anxiety level, can alter the scaling of the BOLD response for the same changes in CBF and $CMRO_2$. In addition, variations in the $CMRO_2$/CBF coupling ratio also strongly affect the magnitude of the BOLD response. In short, the complexity of the BOLD response means that it is difficult to attach a quantitative interpretation to BOLD measurements alone. The solution to this problem is a multi-modal approach combining measurements from other techniques in addition to BOLD. In particular, ASL measurements of CBF significantly expand the potential for interpreting the BOLD response.

In summary, the physics underlying the BOLD effect is reasonably well understood, and much of the effort now is focused on exploiting this understanding to develop new methods for quantitative assessment of brain physiology, particularly oxygen metabolism and cerebral blood volume (especially venous blood volume). If these efforts are successful it will make possible a much more quantitative assessment of human brain physiology.

Acknowledgments

The author has benefitted from helpful discussions of this material with Nic Blockley, Valerie Griffeth and Aaron Simon. This work was supported by grants from the National Institutes of Health: NS-36722, NS-081405 and EB-00790.

References

1. Ogawa S, et al. Brain magnetic resonance imaging with contrast dependent on blood oxygenation. Proc. Natl Acad. Sci. USA. 1990; 87:9868–9872. [PubMed: 2124706]
2. Ogawa S, et al. Intrinsic signal changes accompanying sensory stimulation: functional brain mapping with magnetic resonance imaging. Proc. Natl Acad. Sci. USA. 1992; 89:5951–5955. [PubMed: 1631079]
3. Kwong KK, et al. Dynamic magnetic resonance imaging of human brain activity during primary sensory stimulation. Proc. Natl Acad. Sci. USA. 1992; 89:5675–5679. [PubMed: 1608978]
4. Bandettini PA. Twenty years of functional MRI: the science and the stories. Neuroimage. 2012; 62:575–588. [PubMed: 22542637]
5. Bandettini PA. Functional MRI: a confluence of fortunate circumstances. Neuroimage. 2012; 61:A3–A11.
6. Kwong KK. Record of a single fMRI experiment in May of 1991. Neuroimage. 2012; 62:610–612. [PubMed: 21839841]
7. Ogawa S. Finding the BOLD effect in brain images. Neuroimage. 2012; 62:608–609. [PubMed: 22309802]
8. Rosen BR, Savoy RL. fMRI at 20: has it changed the world? Neuroimage. 2012; 62:1316–1324. [PubMed: 22433659]
9. Turner R. The NIH experience in first advancing fMRI. Neuroimage. 2012; 62:632–636. [PubMed: 21855641]
10. Ugurbil K. Development of functional imaging in the human brain (fMRI); the University of Minnesota experience. Neuroimage. 2012; 62:613–619. [PubMed: 22342875]
11. Pauling L, Coryell CD. The magnetic properties and structure of hemoglobin, oxyhemoglobin, and carbonmonoxyhemoglobin. Proc. Natl Acad. Sci. USA. 1936; 22:210–216. [PubMed: 16577697]
12. Thulborn KR, et al. Oxygenation dependence of the transverse relaxation time of water protons in whole blood at high field. Biochim. Biophys. Acta. 1982; 714:265–270. [PubMed: 6275909]
13. Fox PT, Raichle ME. Focal physiological uncoupling of cerebral blood flow and oxidative metabolism during somatosensory stimulation in human subjects. Proc. Natl Acad. Sci. USA. 1986; 83:1140–1144. [PubMed: 3485282]
14. Detre JA, et al. Perfusion imaging. Magn. Reson. Med. 1992; 23:37–45. [PubMed: 1734182]
15. Mandeville JB, et al. Dynamic functional imaging of relative cerebral blood volume during rat forepaw stimulation. Magn. Reson. Med. 1998; 39:615–624. [PubMed: 9543424]
16. Villringer A, et al. Dynamic imaging with lanthanide chelates in normal brain: contrast due to magnetic susceptibility effects. Magn. Reson. Med. 1988; 6:164–174. [PubMed: 3367774]
17. Lu H, et al. Functional magnetic resonance imaging based on changes in vascular space occupancy. Magn. Reson. Med. 2003; 50:263–274. [PubMed: 12876702]
18. Zhu XH, et al. Advanced *in vivo* heteronuclear MRS approaches for studying brain bioenergetics driven by mitochondria. Methods Mol. Biol. 2009; 489:317–357. [PubMed: 18839099]

19. Buxton RB. Interpreting oxygenation-based neuroimaging signals: the importance and the challenge of understanding brain oxygen metabolism. Front. Neuroenerg. 2010; 2:8.
20. Mintun MA, et al. Brain oxygen utilization measured with O-15 radiotracers and positron emission tomography. J. Nucl. Med. 1984; 25:177–187. [PubMed: 6610032]
21. Griffeth VE, Perthen JE, Buxton RB. Prospects for quantitative fMRI: Investigating the effects of caffeine on baseline oxygen metabolism and the response to a visual stimulus in humans. Neuroimage. 2011; 57:809–816. [PubMed: 21586328]
22. Glover GH. Spiral imaging in fMRI. Neuroimage. 2012; 62:706–712. [PubMed: 22036995]
23. Mansfield P. Multi-planar image formation using NMR spin echoes. J. Phys. C: Solid state Phys. 1977; 10:L55–L58.
24. Van Essen DC, Ugurbil K. The future of the human connectome. Neuroimage. 2012; 62:1299–1310. [PubMed: 22245355]
25. Moeller S, et al. Multiband multislice GE-EPI at 7 tesla, with 16-fold acceleration using partial parallel imaging with application to high spatial and temporal whole-brain fMRI. Magn. Reson. Med. 2010; 63:1144–1153. [PubMed: 20432285]
26. Feinberg DA, et al. Multiplexed echo planar imaging for sub-second whole brain FMRI and fast diffusion imaging. PLoS One. 2010; 5:e15710. [PubMed: 21187930]
27. Van Essen DC, et al. The Human Connectome Project: a data acquisition perspective. Neuroimage. 2012; 62:2222–2231. [PubMed: 22366334]
28. Bandettini PA, et al. Ultrahigh field systems and applications at 7 T and beyond: progress, pitfalls, and potential. Magn. Reson. Med. 2012; 67:317–321. [PubMed: 22083719]
29. Olman CA, et al. Layer-specific fMRI reflects different neuronal computations at different depths in human V1. PLoS One. 2012; 7:e32536. [PubMed: 22448223]
30. Rakic P. Confusing cortical columns. Proc. Natl Acad. Sci. USA. 2008; 105:12099–12100. [PubMed: 18715998]
31. Yacoub E, Harel N, Ugurbil K. High-field fMRI unveils orientation columns in humans. Proc. Natl Acad. Sci. USA. 2008; 105:10607–10612. [PubMed: 18641121]
32. Zhang N, et al. Functional MRI mapping neuronal inhibition and excitation at columnar level in human visual cortex. Exp. Brain Res. 2010; 204:515–524. [PubMed: 20571785]
33. Bandettini PA, et al. Processing strategies for time-course data sets in functional MRI of the human brain. Magn. Reson. Med. 1993; 30:161–173. [PubMed: 8366797]
34. Friston KJ, Jezzard P, Turner R. Analysis of functional MRI time-series. Hum. Brain Mapp. 1994; 1:153–171.
35. Friston KJ, et al. Statistical parametric maps in functional imaging: A general linear approach. Hum. Brain Mapp. 1995; 2:189–210.
36. Smith SM. Overview of fMRI analysis. Br. J. Radiol. 2004; 77:S167–S175. [PubMed: 15677358]
37. Liu TT. The development of event-related fMRI designs. Neuroimage. 2012; 62:1157–1162. [PubMed: 22037002]
38. Buckner RL, et al. Detection of cortical activation during averaged single trials of a cognitive task using functional magnetic resonance imaging. Proc. Natl Acad. Sci. 1996; 93:14878–14883. [PubMed: 8962149]
39. Buckner RL. Event-related fMRI and the hemodynamic response. Hum. Brain Mapp. 1998; 6:373–377. [PubMed: 9788075]
40. Dale AM. Optimal experimental design for event-related fMRI. Hum. Brain Mapp. 1999; 8:109–114. [PubMed: 10524601]
41. Buracas GT, Boynton GM. Efficient design of event-related fMRI experiments using M-sequences. Neuroimage. 2002; 16:801–813. [PubMed: 12169264]
42. Smith SM. The future of FMRI connectivity. Neuroimage. 2012; 62:1257–1266. [PubMed: 22248579]
43. Friston KJ. Functional and effective connectivity: a review. Brain Connect. 2011; 1:13–36. [PubMed: 22432952]
44. Biswal B, et al. Functional connectivity in the motor cortex of resting human brain using echo planar MRI. Magn. Reson. Med. 1995; 34:537–541. [PubMed: 8524021]

45. Fox MD, Raichle ME. Spontaneous fluctuations in brain activity observed with functional magnetic resonance imaging. Nature Rev. Neurosci. 2007; 8:700–711. [PubMed: 17704812]
46. van den Heuvel MP, Hulshoff Pol HE. Exploring the brain network: a review on resting-state fMRI functional connectivity. Eur. Neuropsychopharmacol. 2010; 20:519–534. [PubMed: 20471808]
47. Smith SM, et al. Temporally-independent functional modes of spontaneous brain activity. Proc. Natl Acad. Sci. USA. 2012; 109:3131–3136. [PubMed: 22323591]
48. Handwerker DA, et al. Periodic changes in fMRI connectivity. Neuroimage. 2012; 63:1712–1719. [PubMed: 22796990]
49. Chang C, Glover GH. Time–frequency dynamics of resting-state brain connectivity measured with fMRI. Neuroimage. 2010; 50:81–98. [PubMed: 20006716]
50. Naselaris T, et al. Encoding and decoding in fMRI. Neuroimage. 2011; 56:400–410. [PubMed: 20691790]
51. Tong F, Pratte MS. Decoding patterns of human brain activity. Annu. Rev. Psychol. 2012; 63:483–509. [PubMed: 21943172]
52. Mori S, Zhang J. Principles of diffusion tensor imaging and its applications to basic neuroscience research. Neuron. 2006; 51:527–539. [PubMed: 16950152]
53. Stejskal EO, Tanner JE. Spin-diffusion measurements: spin echoes in the presence of a time-dependent field gradient. J. Chem. Phys. 1965; 42:288–292.
54. Hsu EW, Mori S. Analytical expressions for the NMR apparent diffusion coefficients in an anisotropic system and a simplified method for determining fiber orientation. Magn. Reson. Med. 1995; 34:194–200. [PubMed: 7476078]
55. Basser PJ, Jones DK. Diffusion-tensor MRI: theory, experimental design and data analysis—a technical review. NMR Biomed. 2002; 15:456–467. [PubMed: 12489095]
56. Behrens TE, Johansen-Berg H. Relating connectional architecture to grey matter function using diffusion imaging. Phil. Trans. R. Soc. Lond. B. 2005; 360:903–911. [PubMed: 16087435]
57. Johansen-Berg H, Rushworth MF. Using diffusion imaging to study human connectional anatomy. Annu. Rev. Neurosci. 2009; 32:75–94. [PubMed: 19400718]
58. Tuch, DS., et al. High angular resolution diffusion imaging of the human brain. 7th Meeting of the International Society for Magnetic Resonance in Medicine; Philadelphia, PA. 1999.
59. Frank LR. Characterization of anisotropy in high angular resolution diffusion-weighted MRI. Magn. Reson. Med. 2002; 47:1083–1099. [PubMed: 12111955]
60. Behrens TE, et al. Probabilistic diffusion tractography with multiple fibre orientations: what can we gain? Neuroimage. 2007; 34:144–155. [PubMed: 17070705]
61. Dell'Acqua F, Catani M. Structural human brain networks: hot topics in diffusion tractography. Curr. Opin. Neurol. 2012; 25:375–383. [PubMed: 22766720]
62. Small SA, et al. A pathophysiological framework of hippocampal dysfunction in ageing and disease. Nature Rev. Neurosci. 2011; 12:585–601. [PubMed: 21897434]
63. Brown GG, Clark C, Liu TT. Measurement of cerebral perfusion with arterial spin labeling: II. Applications. J. Int. Neuropsychol. Soc. 2007; 13:526–538. [PubMed: 17445302]
64. Liu TT, Brown GG. Measurement of cerebral perfusion with arterial spin labeling: I. Methods. J. Int. Neuropsychol. Soc. 2007; 13:517–525. [PubMed: 17445301]
65. Petersen ET, Lim T, Golay X. Model-free arterial spin labeling quantification approach for perfusion MRI. Magn. Reson. Med. 2006; 55:219–232. [PubMed: 16416430]
66. Qiu D, et al. CBF measurements using multidelay pseudocontinuous and velocity-selective arterial spin labeling in patients with long arterial transit delays: comparison with xenon CT CBF. J. Magn. Reson. Imag. 2012; 36:110–119.
67. MacIntosh BJ, et al. Assessment of arterial arrival times derived from multiple inversion time pulsed arterial spin labeling MRI. Magn. Reson. Med. 2010; 63:641–647. [PubMed: 20146233]
68. Alsop DC, Detre JA. Reduced transit-time sensitivity in noninvasive magnetic resonance imaging of human cerebral blood flow. J. Cereb. Blood Flow Metab. 1996; 16:1236–1249. [PubMed: 8898697]
69. Buxton RB, et al. A general kinetic model for quantitative perfusion imaging with arterial spin labeling. Magn. Reson. Med. 1998; 40:383–396. [PubMed: 9727941]

70. Wong EC, Buxton RB, Frank LR. Quantitative imaging of perfusion using a single subtraction (QUIPSS and QUIPSS II). Magn. Reson. Med. 1998; 39:702–708. [PubMed: 9581600]
71. Buerk D, Bridges E. A simplified algorithm for computing the variation in oxyhemoglobin saturation with pH pCO_2, T and DPG. Chem. Eng. Commun. 1986; 47:113–124.
72. Devor A, et al. 'Overshoot' of o2 is required to maintain baseline tissue oxygenation at locations distal to blood vessels. J. Neurosci. 2011; 31:13676–13681. [PubMed: 21940458]
73. Gobel U, Theilen H, Kuschinsky W. Congruence of total and perfused capillary network in rat brains. Circ. Res. 1990; 66:271–281. [PubMed: 1688741]
74. Attwell D, Iadecola C. The neural basis of functional brain imaging signals. Trends Neurosci. 2002; 25:621–625. [PubMed: 12446129]
75. Nicholls, DG.; Ferguson, FJ. Bioenergetics. Vol. 3. San Diego, CA: Academic Press; 2002.
76. Erecinska M, Silver IA. Ions and energy in mammalian brain. Prog. Neurobiol. 1994; 43:37–71. [PubMed: 7972852]
77. Buxton, RB. Introduction to Functional Magnetic Resonance Imaging: Principles and Techniques. Cambridge: Cambridge University Press; 2009.
78. Blaustein MP. Calcium transport and buffering in neurons. Trends Neurosci. 1988; 11:438–443. [PubMed: 2469161]
79. Ames A III. CNS energy metabolism as related to function. Brain Res. Rev. 2000; 34:42–68. [PubMed: 11086186]
80. Attwell D, Laughlin SB. An energy budget for signaling in the grey matter of the brain. J. Cereb. Blood Flow Metab. 2001; 21:1133–1145. [PubMed: 11598490]
81. Laughlin SB, Sejnowski TJ. Communication in neuronal networks. Science. 2003; 301:1870–1874. [PubMed: 14512617]
82. Logothetis NK. The neural basis of the blood-oxygen-level-dependent functional magnetic resonance imaging signal. Phil. Trans. R. Soc. Lond. B. 2002; 357:1003–1037. [PubMed: 12217171]
83. Siesjo, B. Brain Energy Metabolism. New York: Wiley; 1978.
84. Fox PT, et al. Nonoxidative glucose consumption during focal physiologic neural activity. Science. 1988; 241:462–464. [PubMed: 3260686]
85. Pellerin L, et al. Activity-dependent regulation of energy metabolism by astrocytes: an update. Glia. 2007; 55:1251–1262. [PubMed: 17659524]
86. Lin AL, et al. Nonlinear coupling between cerebral blood flow, oxygen consumption, and ATP production in human visual cortex. Proc. Natl Acad. Sci. USA. 2010; 107:8446–8451. [PubMed: 20404151]
87. Villringer A, Dirnagl U. Coupling of brain activity and cerebral blood flow: basis of functional neuroimaging. Cerebrovasc. Brain Metab. Rev. 1995; 7:240–276. [PubMed: 8519605]
88. Girouard H, Iadecola C. Neurovascular coupling in the normal brain and in hypertension, stroke, and Alzheimer disease. J. Appl. Physiol. 2006; 100:328–335. [PubMed: 16357086]
89. Hamel E. Perivascular nerves and the regulation of cerebrovascular tone. J. Appl. Physiol. 2006; 100:1059–1064. [PubMed: 16467392]
90. Koehler RC, Roman RJ, Harder DR. Astrocytes and the regulation of cerebral blood flow. Trends Neurosci. 2009; 32:160–169. [PubMed: 19162338]
91. Petzold GC, Murthy VN. Role of astrocytes in neurovascular coupling. Neuron. 2011; 71:782–797. [PubMed: 21903073]
92. Faraci FM, Sobey CG. Role of potassium channels in regulation of cerebral vascular tone. J. Cereb. Blood Flow Metab. 1998; 18:1047–1063. [PubMed: 9778181]
93. Lassen, NA. Cations as mediators of functional hyperemia in the brain. In: Lassen, NA., et al., editors. Brain Work and Mental Activity: Quantitative Studies with Radioactive Tracers. Copenhagen: Munksgaard; 1991. p. 68-77.
94. Filosa JA, et al. Local potassium signaling couples neuronal activity to vasodilation in the brain. Nature Neurosci. 2006; 9:1397–1403. [PubMed: 17013381]
95. Dirnagl U, et al. Coupling of cerebral blood flow to neuronal activation: role of adenosine and nitric oxide. Am. J. Physiol. 1994; 267:H296–H301. [PubMed: 8048594]

96. Winn, HR.; Ngai, AC.; Ko, KR. Role of adenosine in regulating microvascular CBF in activated sensory cortex. In: Lassen, NA., et al., editors. Brain Work and Mental Activity: Quantitative Studies with Radioactive Tracers. Copenhagen: Munksgaard; 1991.

97. Haydon PG, Carmignoto G. Astrocyte control of synaptic transmission and neurovascular coupling. Physiol. Rev. 2006; 86:1009–1031. [PubMed: 16816144]

98. Fredholm BB, et al. Actions of caffeine in the brain with special reference to factors that contribute to it widespread use. Pharmcol. Rev. 1999; 51:83–133.

99. Perthen JE, et al. Caffeine-induced uncoupling of cerebral blood flow and oxygen metabolism: a calibrated BOLD fMRI study. Neuroimage. 2008; 40:237–247. [PubMed: 18191583]

100. Straub SV, Nelson MT. Astrocytic calcium signaling: the information currency coupling neuronal activity to the cerebral microcirculation. Trends Cardiovasc. Med. 2007; 17:183–190. [PubMed: 17662912]

101. Cauli B, et al. Cortical GABA interneurons in neurovascular coupling: relays for subcortical vasoactive pathways. J. Neurosci. 2004; 24:8940–8949. [PubMed: 15483113]

102. Belanger M, Allaman I, Magistretti PJ. Brain energy metabolism: focus on astrocyte-neuron metabolic cooperation. Cell Metab. 2011; 14:724–738. [PubMed: 22152301]

103. Nedergaard M, Ransom B, Goldman SA. New roles for astrocytes: redefining the functional architecture of the brain. Trends Neurosci. 2003; 26:523–530. [PubMed: 14522144]

104. Iadecola C, Nedergaard M. Glial regulation of the cerebral microvasculature. Nature Neurosci. 2007; 10:1369–1376. [PubMed: 17965657]

105. Andresen J, Shafi NI, Bryan RM Jr. Endothelial influences on cerebrovascular tone. J. Appl. Physiol. 2006; 100:318–327. [PubMed: 16357085]

106. Lecrux C, Hamel E. The neurovascular unit in brain function and disease. Acta Physiol. 2011; 203:47–59.

107. Estrada C, DeFelipe J. Nitric oxide-producing neurons in the neocortex: morphological and functional relationship with intraparenchymal microvasculature. Cereb. Cortex. 1998; 8:193–203. [PubMed: 9617914]

108. Pike GB. Quantitative functional MRI: concepts, issues and future challenges. Neuroimage. 2012; 62:1234–1240. [PubMed: 22056462]

109. Blockley NP, et al. A review of calibrated blood oxygenation level-dependent (BOLD) methods for the measurement of task-induced changes in brain oxygen metabolism. NMR Biomed. 2012 at press.

110. Yablonskiy DA, Sukstanskii AL, He X. Blood oxygenation level-dependent (BOLD)-based techniques for the quantification of brain hemodynamic and metabolic properties—theoretical models and experimental approaches. NMR Biomed. 2012

111. Miller KL, et al. Functional brain imaging using a blood oxygenation sensitive steady state. Magn. Reson. Med. 2003; 50:675–683. [PubMed: 14523951]

112. Miller KL. FMRI using balanced steady-state free precession (SSFP). Neuroimage. 2012; 62:713–719. [PubMed: 22036996]

113. Griffeth VE, Buxton RB. A theoretical framework for estimating cerebral oxygen metabolism changes using the calibrated-BOLD method: modeling the effects of blood volume distribution, hematocrit, oxygen extraction fraction, and tissue signal properties on the BOLD signal. Neuroimage. 2011; 58:198–212. [PubMed: 21669292]

114. Uludag K, Muller-Bierl B, Ugurbil K. An integrative model for neuronal activity-induced signal changes for gradient and spin echo functional imaging. Neuroimage. 2009; 48:150–165. [PubMed: 19481163]

115. Spees WM, et al. Water proton MR properties of human blood at 1.5 Tesla: magnetic susceptibility, T(1), T(2), T*(2), and non-Lorentzian signal behavior. Magn. Reson. Med. 2001; 45:533–542. [PubMed: 11283978]

116. Ogawa S, et al. Functional brain mapping by blood oxygenation level-dependent contrast magnetic resonance imaging: a comparison of signal characteristics with a biophysical model. Biophys. J. 1993; 64:803–812. [PubMed: 8386018]

117. Dickson JD, et al. Quantitative BOLD: the effect of diffusion. J. Magn. Reson. Imag. 2010; 32:953–961.

118. Lauwers F, et al. Morphometry of the human cerebral cortex microcirculation: general characteristics and space-related profiles. Neuroimage. 2008; 39:936–948. [PubMed: 17997329]
119. Lorthois S, Cassot F, Lauwers F. Simulation study of brain blood flow regulation by intra-cortical arterioles in an anatomically accurate large human vascular network: II. Flow variations induced by global or localized modifications of arteriolar diameters. Neuroimage. 2011; 54:2840–2853. [PubMed: 21047557]
120. Lorthois S, Cassot F, Lauwers F. Simulation study of brain blood flow regulation by intra-cortical arterioles in an anatomically accurate large human vascular network: I. Methodology and baseline flow. Neuroimage. 2011; 54:1031–1042. [PubMed: 20869450]
121. Hirsch S, et al. Topology and hemodynamics of the cortical cerebrovascular system. J. Cereb. Blood Flow Metab. 2012; 32:952–967. [PubMed: 22472613]
122. Fang Q, et al. Oxygen advection and diffusion in a three-dimensional vascular anatomical network. Opt. Express. 2008; 16:17530–17541. [PubMed: 18958033]
123. Boas DA, et al. A vascular anatomical network model of the spatio-temporal response to brain activation. Neuroimage. 2008; 40:1116–1129. [PubMed: 18289880]
124. Yablonskiy DA, Haacke EM. Theory of NMR signal behavior in magnetically inhomogenous tissues: the static dephasing regime. Magn. Reson. Med. 1994; 32:749–763. [PubMed: 7869897]
125. Kiselev VG, Posse S. Analytical model of susceptibility-induced MR signal dephasing: effect of diffusion in a microvascular network. Magn. Reson. Med. 1999; 41:499–509. [PubMed: 10204873]
126. Sukstanskii AL, Yablonskiy DA. Gaussian approximation in the theory of MR signal formation in the presence of structure-specific magnetic field inhomogeneities. J. Magn. Reson. 2003; 163:236–247. [PubMed: 12914839]
127. Bauer WR, et al. Theory of the BOLD effect in the capillary region: an analytical approach for the determination of T2 in the capillary network of myocardium. Magn. Reson. Med. 1999; 41:51–62. [PubMed: 10025611]
128. Stables LA, Kennan RP, Gore JC. Asymmetric spin-echo imaging of magnetically inhomogeneous systems: theory, experiment, and numerical studies. Magn. Reson. Med. 1998; 40:432–442. [PubMed: 9727947]
129. Boxerman, JL., et al. MR contrast due to microscopically heterogeneous magnetic susceptibility: cylindrical geometry. Proc. SMRM, 12th Annual Meeting; New York, NY. 1993.
130. Davis TL, et al. Calibrated functional MRI: mapping the dynamics of oxidative metabolism. Proc. Natl Acad. Sci. USA. 1998; 95:1834–1839. [PubMed: 9465103]
131. Dickson JD, et al. Quantitative phenomenological model of the BOLD contrast mechanism. J. Magn. Reson. 2011; 212:17–25. [PubMed: 21782488]
132. Menon RS. The great brain versus vein debate. Neuroimage. 2012; 62:970–974. [PubMed: 21939776]
133. Lai S, et al. Identification of vascular structures as a major source of signal contrast in high resolution 2D and 3D functional activation imaging of the motor cortex at 1.5 T: preliminary results. Magn. Reson. Med. 1993; 30:387–392. [PubMed: 8412613]
134. Pflugfelder D, et al. On the numerically predicted spatial BOLD fMRI specificity for spin echo sequences. Magn. Reson. Imag. 2011; 29:1195–1204.
135. Yacoub E, et al. Spin-echo fMRI in humans using high spatial resolutions and high magnetic fields. Magn. Reson. Med. 2003; 49:655–664. [PubMed: 12652536]
136. van Zijl PC, et al. Quantitative assessment of blood flow, blood volume and blood oxygenation effects in functional magnetic resonance imaging. Nature Med. 1998; 4:159–167. [PubMed: 9461188]
137. Oja JME, et al. Venous blood effects in spin-echo fMRI of human brain. Magn. Reson. Med. 1999; 42:617–626. [PubMed: 10502748]
138. Lee SP, et al. Diffusion-weighted spin-echo fMRI at 9.4 T: microvascular/tissue contribution to BOLD signal changes. Magn. Reson. Med. 1999; 42:919–928. [PubMed: 10542351]
139. Kiselev VG, et al. Vessel size imaging in humans. Magn. Reson. Med. 2005; 53:553–563. [PubMed: 15723391]

140. Tropres I, et al. Vessel size imaging. Magn. Reson. Med. 2001; 45:397–408. [PubMed: 11241696]
141. Dennie J, et al. NMR imaging of changes in vascular morphology due to tumor angiogenesis. Magn. Reson. Med. 1998; 40:793–799. [PubMed: 9840821]
142. Boxerman JL, et al. The intravascular contribution to fMRI signal change: Monte Carlo modeling and diffusion-weighted studies in vivo. Magn. Reson. Med. 1995; 34:4–10. [PubMed: 7674897]
143. Donahue MJ, et al. Blood oxygenation level-dependent (BOLD) total and extravascular signal changes and Delta R2* in human visual cortex at 1.5, 3.0 and 7.0 T. NMR Biomed. 2011; 24:25–34. [PubMed: 21259367]
144. Donahue MJ, et al. Cerebral blood flow, blood volume, and oxygen metabolism dynamics in human visual and motor cortex as measured by whole-brain multi-modal magnetic resonance imaging. J. Cereb. Blood Flow Metab. 2009; 29:1856–1866. [PubMed: 19654592]
145. Zhao JM, et al. Oxygenation and hematocrit dependence of transverse relaxation rates of blood at 3 T. Magn. Reson. Med. 2007; 58:592–597. [PubMed: 17763354]
146. Buxton RB, et al. Modeling the hemodynamic response to brain activation. Neuroimage. 2004; 23:S220–S233. [PubMed: 15501093]
147. Grubb RL, et al. The effects of changes in $PaCO_2$ on cerebral blood volume, blood flow, and vascular mean transit time. Stroke. 1974; 5:630–639. [PubMed: 4472361]
148. Chen JJ, Pike GB. BOLD-specific cerebral blood volume and blood flow changes during neuronal activation in humans. NMR Biomed. 2009; 22:1054–1062. [PubMed: 19598180]
149. Hillman EM, et al. Depth-resolved optical imaging and microscopy of vascular compartment dynamics during somatosensory stimulation. Neuroimage. 2007; 35:89–104. [PubMed: 17222567]
150. Kim T, et al. Arterial versus total blood volume changes during neural activity-induced cerebral blood flow change: implication for BOLD fMRI. J. Cereb. Blood Flow Metab. 2007; 27:1235–1247. [PubMed: 17180136]
151. Kida I, Rothman DL, Hyder F. Dynamics of changes in blood flow, volume, and oxygenation: implications for dynamic functional magnetic resonance imaging calibration. J. Cereb. Blood Flow Metab. 2007; 27:690–696. [PubMed: 17033688]
152. Mark CI, Pike GB. Indication of BOLD-specific venous flow-volume changes from precisely controlled hyperoxic versus hypercapnic calibration. J. Cereb. Blood Flow Metab. 2012; 32:709–719. [PubMed: 22167238]
153. Mark CI, Fisher JA, Pike GB. Improved fMRI calibration: precisely controlled hyperoxic versus hypercapnic stimuli. Neuroimage. 2011; 54:1102–1111. [PubMed: 20828623]
154. Buxton RB, Wong EC, Frank LR. Dynamics of blood flow and oxygenation changes during brain activation: the balloon model. Magn. Reson. Med. 1998; 39:855–864. [PubMed: 9621908]
155. Obata T, et al. Discrepancies between BOLD and flow dynamics in primary and supplementary motor areas: application of the balloon model to the interpretation of BOLD transients. Neuroimage. 2004; 21:144–153. [PubMed: 14741651]
156. Gauthier CJ, et al. Elimination of visually evoked BOLD responses during carbogen inhalation: implications for calibrated MRI. Neuroimage. 2011; 54:1001–1011. [PubMed: 20887792]
157. Hoge RD. Calibrated FMRI. Neuroimage. 2012; 62:930–937. [PubMed: 22369993]
158. Xu F, et al. The influence of carbon dioxide on brain activity and metabolism in conscious humans. J. Cereb. Blood Flow Metab. 2011; 31:58–67. [PubMed: 20842164]
159. Zappe AC, et al. The influence of moderate hypercapnia on neural activity in the anesthetized nonhuman primate. Cereb. Cortex. 2008; 18:2666–2673. [PubMed: 18326521]
160. Jain V, et al. Rapid magnetic resonance measurement of global cerebral metabolic rate of oxygen consumption in humans during rest and hypercapnia. J. Cereb. Blood Flow Metab. 2011; 31:1504–1512. [PubMed: 21505481]
161. Thesen T, et al. Depression of cortical activity in humans by mild hypercapnia. Hum. Brain Mapp. 2012; 33:715–726. [PubMed: 21500313]

162. Chen JJ, Pike GB. Global cerebral oxidative metabolism during hypercapnia and hypocapnia in humans: implications for BOLD fMRI. J. Cereb. Blood Flow Metab. 2010; 30:1094–1099. [PubMed: 20372169]
163. Chiarelli PA, et al. A calibration method for quantitative BOLD fMRI based on hyperoxia. Neuroimage. 2007; 37:808–820. [PubMed: 17632016]
164. Blockley NP, Griffeth VE, Buxton RB. A general analysis of calibrated BOLD methodology for measuring CMRO2 responses: comparison of a new approach with existing methods. Neuroimage. 2012; 60:279–289. [PubMed: 22155329]
165. Xu F, et al. Effect of hypoxia and hyperoxia on cerebral blood flow, blood oxygenation, and oxidative metabolism. J. Cereb. Blood Flow Metab. 2012; 32:1909–1918. [PubMed: 22739621]
166. Christen T, et al. Evaluation of a quantitative blood oxygenation level-dependent (qBOLD) approach to map local blood oxygen saturation. NMR Biomed. 2011; 24:393–403. [PubMed: 20960585]
167. He X, Yablonskiy DA. Quantitative BOLD: mapping of human cerebral deoxygenated blood volume and oxygen extraction fraction: default state. Magn. Reson. Med. 2007; 57:115–126. [PubMed: 17191227]
168. Driver ID, et al. Calibrated BOLD using direct measurement of changes in venous oxygenation. Neuroimage. 2012; 63:1178–1187. [PubMed: 22971549]
169. Lin AL, et al. Evaluation of MRI models in the measurement of CMRO2 and its relationship with CBF. Magn. Reson. Med. 2008; 60:380–389. [PubMed: 18666102]
170. Yucel MA, et al. Calibrating the BOLD signal during a motor task using an extended fusion model incorporating DOT, BOLD and ASL data. Neuroimage. 2012; 61:1268–1276. [PubMed: 22546318]
171. Fleisher AS, et al. Cerebral perfusion and oxygenation differences in Alzheimer's disease risk. Neurobiol. Aging. 2009; 30:1737–1748. [PubMed: 18325636]
172. Chen Y, Parrish TB. Caffeine dose effect on activation-induced BOLD and CBF responses. Neuroimage. 2009; 46:577–583. [PubMed: 19289172]
173. Yesilyurt B, et al. Relationship of the BOLD signal with VEP for ultrashort duration visual stimuli (0.1 to 5 ms) in humans. J. Cereb. Blood Flow Metab. 2010; 30:449–458. [PubMed: 19844243]
174. Cohen ER, Ugurbil K, Kim SG. Effect of basal conditions on the magnitude and dynamics of the blood oxygenation level-dependent fMRI response. J. Cereb. Blood Flow Metab. 2002; 22:1042–1053. [PubMed: 12218410]
175. Liu TT, et al. Caffeine alters the temporal dynamics of the visual BOLD response. Neuroimage. 2004; 23:1402–1413. [PubMed: 15589104]
176. Behzadi Y, Liu TT. Caffeine reduces the initial dip in the visual BOLD response at 3 T. Neuroimage. 2006; 32:9–15. [PubMed: 16635577]
177. Behzadi Y, Liu TT. An arteriolar compliance model of the cerebral blood flow response to neural stimulus. Neuroimage. 2005; 25:1100–1111. [PubMed: 15850728]
178. Hu X, Yacoub E. The story of the initial dip in fMRI. Neuroimage. 2012; 62:1103–1108. [PubMed: 22426348]
179. van Zijl PC, Hua J, Lu H. The BOLD post-stimulus undershoot, one of the most debated issues in fMRI. Neuroimage. 2012; 62:1092–1102. [PubMed: 22248572]
180. Buxton RB. Dynamic models of BOLD contrast. Neuroimage. 2012
181. Frostig RD, et al. Cortical functional architecture and local coupling between neuronal activity and the microcirculation revealed by *in vivo* high-resolution optical imaging of intrinsic signals. Proc. Natl Acad. Sci. USA. 1990; 87:6082–6086. [PubMed: 2117272]
182. Grinvald A, et al. High-resolution optical imaging of functional brain architecture in the awake monkey. Proc. Natl Acad. Sci. USA. 1991; 88:11559–11563. [PubMed: 1763070]
183. Menon RS, et al. BOLD based functional MRI at 4 tesla includes a capillary bed contribution: echo-planar imaging correlates with previous optical imaging using intrinsic signals. Magn. Reson. Med. 1995; 33:453–459. [PubMed: 7760717]

184. Devor A, et al. Coupling of total hemoglobin concentration, oxygenation, and neural activity in rat somatosensory cortex. Neuron. 2003; 39:353–359. [PubMed: 12873390]

185. Sirotin YB, et al. Spatiotemporal precision and hemodynamic mechanism of optical point spreads in alert primates. Proc. Natl Acad. Sci. USA. 2009; 106:18390–18395. [PubMed: 19828443]

186. Uludag K. To dip or not to dip: reconciling optical imaging and fMRI data. Proc. Natl Acad. Sci. USA. 2010; 107:E23. (author reply E24). [PubMed: 20142469]

187. Chen BR, et al. High-speed vascular dynamics of the hemodynamic response. Neuroimage. 2011; 54:1021–1030. [PubMed: 20858545]

188. Yacoub E, et al. Investigation of the initial dip in fMRI at 7 Tesla. NMR Biomed. 2001; 14:408–412. [PubMed: 11746933]

189. Sadaghiani S, Ugurbil K, Uludag K. Neural activity-induced modulation of BOLD poststimulus undershoot independent of the positive signal. Magn. Reson. Imag. 2009; 27:1030–1038.

190. Frahm J, et al. Dynamic uncoupling and recoupling of perfusion and oxidative metabolism during focal activation in man. Magn. Reson. Med. 1996; 35:143–148. [PubMed: 8622575]

191. Kruger G, Kleinschmidt A, Frahm J. Dynamic MRI sensitized to cerebral blood oxygenation and flow during sustained activation of human visual cortex. Magn. Reson. Med. 1996; 35:797–800. [PubMed: 8744004]

192. Mandeville JB, et al. Evidence of a cerebrovascular post-arteriole Windkessel with delayed compliance. J. Cereb. Blood Flow Metabol. 1999; 19:679–689.

193. Kim T, Kim SG. Temporal dynamics and spatial specificity of arterial and venous blood volume changes during visual stimulation: implication for BOLD quantification. J. Cereb. Blood Flow Metab. 2011; 31:1211–1222. [PubMed: 21179068]

194. Drew PJ, Shih AY, Kleinfeld D. Fluctuating and sensory-induced vasodynamics in rodent cortex extend arteriole capacity. Proc. Natl Acad. Sci. USA. 2011; 108:8473–8478. [PubMed: 21536897]

195. Dechent P, et al. Basal cerebral blood volume during the poststimulation undershoot in BOLD MRI of the human brain. J. Cereb. Blood Flow Metab. 2011; 31:82–89. [PubMed: 20736964]

196. Frahm J, et al. The post-stimulation undershoot in BOLD fMRI of human brain is not caused by elevated cerebral blood volume. Neuroimage. 2008; 40:473–481. [PubMed: 18201912]

197. Lu H, et al. Sustained poststimulus elevation in cerebral oxygen utilization after vascular recovery. J. Cereb. Blood Flow Metab. 2004; 24:764–770. [PubMed: 15241184]

198. Tuunanen PI, Vidyasagar R, Kauppinen RA. Effects of mild hypoxic hypoxia on poststimulus undershoot of blood-oxygenation-level-dependent fMRI signal in the human visual cortex. Magn. Reson. Imag. 2006; 24:993–999.

199. Schroeter ML, et al. Investigating the post-stimulus undershoot of the BOLD signal—a simultaneous fMRI and fNIRS study. Neuroimage. 2006; 30:349–358. [PubMed: 16257236]

200. Chen JJ, Pike GB. Origins of the BOLD post-stimulus undershoot. Neuroimage. 2009; 46:559–568. [PubMed: 19303450]

201. Donahue MJ, et al. Hemodynamic changes after visual stimulation and breath holding provide evidence for an uncoupling of cerebral blood flow and volume from oxygen metabolism. J. Cereb. Blood Flow Metab. 2009; 29:176–185. [PubMed: 18797471]

202. Hua J, et al. Physiological origin for the BOLD poststimulus undershoot in human brain: vascular compliance versus oxygen metabolism. J. Cereb. Blood Flow Metab. 2011; 31:1599–1611. [PubMed: 21468090]

203. Fuchtemeier M, et al. Elevating intracranial pressure reverses the decrease in deoxygenated hemoglobin and abolishes the post-stimulus overshoot upon somatosensory activation in rats. Neuroimage. 2010; 52:445–454. [PubMed: 20420930]

204. Lu H, Ge Y. Quantitative evaluation of oxygenation in venous vessels using T2-relaxation-under-spin-tagging MRI. Magn. Reson. Med. 2008; 60:357–363. [PubMed: 18666116]

205. Xu F, Ge Y, Lu H. Noninvasive quantification of whole-brain cerebral metabolic rate of oxygen (CMRO(2)) by MRI. Magn. Reson. Med. 2009; 62:141–148. [PubMed: 19353674]

206. Liu P, Xu F, Lu H. Test-retest reproducibility of a rapid method to measure brain oxygen metabolism. Magn. Reson. Med. 2012; 69:675–681. [PubMed: 22517498]

207. Xu F, et al. On improving the speed and reliability of T2-relaxation-under-spin-tagging (TRUST) MRI. Magn. Reson. Med. 2012; 68:198–204. [PubMed: 22127845]

208. Guo J, Wong EC. Venous oxygenation mapping using velocity-selective excitation and arterial nulling. Magn. Reson. Med. 2012; 68:1458–1471. [PubMed: 22294414]

209. Bolar DS, et al. QUantitative Imaging of eXtraction of oxygen and TIssue consumption (QUIXOTIC) using venular-targeted velocity-selective spin labeling. Magn. Reson. Med. 2011; 66:1550–1562. [PubMed: 21674615]

210. Lu H, et al. Calibration and validation of TRUST MRI for the estimation of cerebral blood oxygenation. Magn. Reson. Med. 2012; 67:42–49. [PubMed: 21590721]

211. Haacke EM, et al. *In vivo* measurement of blood oxygen saturation using magnetic resonance imaging: a direct validation of the blood oxygen level-dependent concept in functional brain imaging. Hum. Brain Mapp. 1997; 5:341–346. [PubMed: 20408238]

212. Fernandez-Seara MA, et al. MR susceptometry for measuring global brain oxygen extraction. Magn. Reson. Med. 2006; 55:967–973. [PubMed: 16598726]

213. Jain V, Langham MC, Wehrli FW. MRI estimation of global brain oxygen consumption rate. J. Cereb. Blood Flow Metab. 2010; 30:1598–1607. [PubMed: 20407465]

214. Fan AP, et al. Phase-based regional oxygen metabolism (PROM) using MRI. Magn. Reson. Med. 2012; 67:669–678. [PubMed: 21713981]

215. An H, Lin W. Quantitative measurements of cerebral blood oxygen saturation using magnetic resonance imaging. J. Cereb. Blood Flow Metab. 2000; 20:1225–1236. [PubMed: 10950383]

216. An H, Lin W. Cerebral oxygen extraction fraction and cerebral venous blood volume measurements using MRI: effects of magnetic field variation. Magn. Reson. Med. 2002; 47:958–966. [PubMed: 11979575]

217. Fujita N, et al. Quantitative mapping of cerebral deoxyhemoglobin content using MR imaging. Neuroimage. 2003; 20:2071–2083. [PubMed: 14683711]

218. An H, et al. Evaluation of MR-derived cerebral oxygen metabolic index in experimental hyperoxic hypercapnia, hypoxia, and ischemia. Stroke. 2009; 40:2165–2172. [PubMed: 19359642]

219. Sedlacik J, Reichenbach JR. Validation of quantitative estimation of tissue oxygen extraction fraction and deoxygenated blood volume fraction in phantom and in vivo experiments by using MRI. Magn. Reson. Med. 2010; 63:910–921. [PubMed: 20373392]

220. Sohlin MC, Schad LR. Susceptibility-related MR signal dephasing under nonstatic conditions: experimental verification and consequences for qBOLD measurements. J. Magn. Reson. Imag. 2011; 33:417–425.

221. Wang X, Sukstanskii AL, Yablonskiy DA. Optimization strategies for evaluation of brain hemodynamic parameters with qBOLD technique. Magn. Reson. Med. 2012; 69:1034–1043. [PubMed: 22623013]

222. Christen T, et al. Measuring brain oxygenation in humans using a multiparametric quantitative blood oxygenation level dependent MRI approach. Magn. Reson. Med. 2012; 68:905–911. [PubMed: 22162074]

223. Gauthier CJ, et al. Absolute quantification of resting oxygen metabolism and metabolic reactivity during functional activation using QUO2 MRI. Neuroimage. 2012; 63:1353–1363. [PubMed: 22986357]

224. Gauthier CJ, Hoge RD. A generalized procedure for calibrated MRI incorporating hyperoxia and hypercapnia. Hum. Brain Mapp. 2012; 34:1053–1069. [PubMed: 23015481]

225. Gauthier CJ, Hoge RD. Magnetic resonance imaging of resting OEF and CMRO(2) using a generalized calibration model for hypercapnia and hyperoxia. Neuroimage. 2012; 60:1212–1225. [PubMed: 22227047]

226. Bulte DP, et al. Quantitative measurement of cerebral physiology using respiratory-calibrated MRI. Neuroimage. 2012; 60:582–591. [PubMed: 22209811]

227. Blockley NP, et al. An analysis of the use of hyperoxia for measuring venous cerebral blood volume: comparison of the existing method with a new analysis approach. Neuroimage. 2013; 72:33–40. [PubMed: 23370053]

228. Miller KL, et al. Nonlinear temporal dynamics of the cerebral blood flow response. Hum. Brain Mapp. 2001; 13:1–12. [PubMed: 11284042]

229. Leontiev O, et al. Coupling of cerebral blood flow and oxygen metabolism is conserved for chromatic and luminance stimuli in human visual cortex. Neuroimage. 2013; 68:221–228. [PubMed: 23238435]

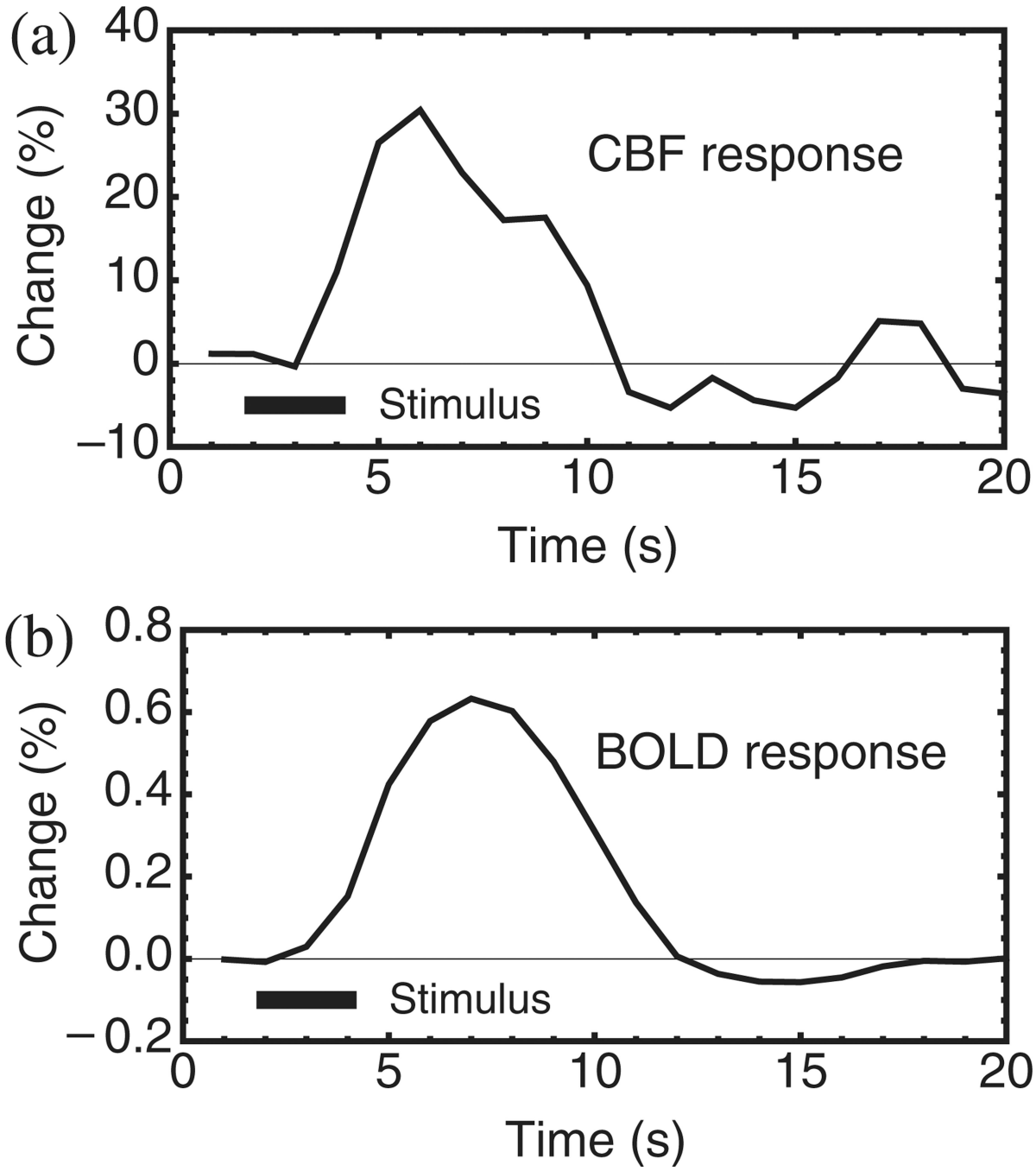

Figure 1.

CBF and BOLD responses in human primary motor cortex to 2 s of finger tapping. (*a*) The brief stimulus evokes a strong change in CBF measured with an ASL method. (*b*) The CBF change is accompanied by an increase in venous blood oxygenation, giving rise to the BOLD response measured with fMRI. Figure reproduced from [19] based on data from [228].

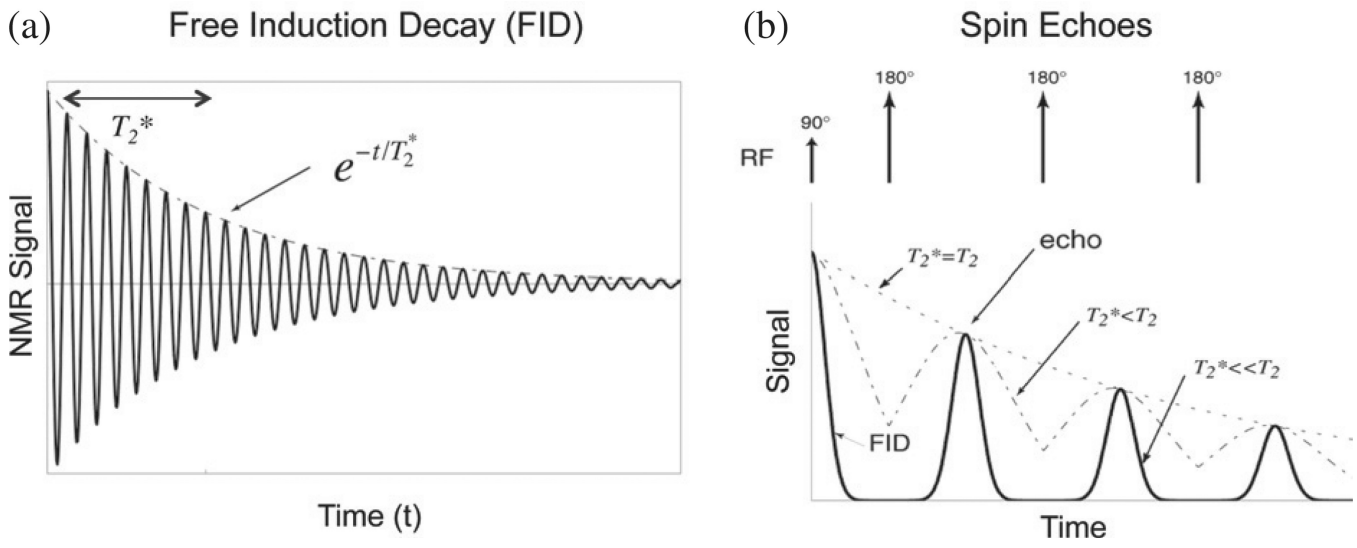

Figure 2.

NMR signal decay. (*a*) After an initial RF pulse creates transverse magnetization, the primary signal measured in fMRI is the FID, called a GE signal in MRI terminology. The signal decays approximately with a time constant T_2^* due to both intrinsic T_2 decay plus dephasing due to magnetic field inhomogeneities within an image voxel. (*b*) Subsequent RF pulses create *spin echoes* of the original signal, reversing the effects of the magnetic field inhomogeneities. The signal at each SE decays with time constant T_2 ($>T_2^*$) on subsequent echoes. The primary origin of the BOLD effect in fMRI is that blood oxygenation affects T_2^*. Figure adapted from [77] with permission of the author.

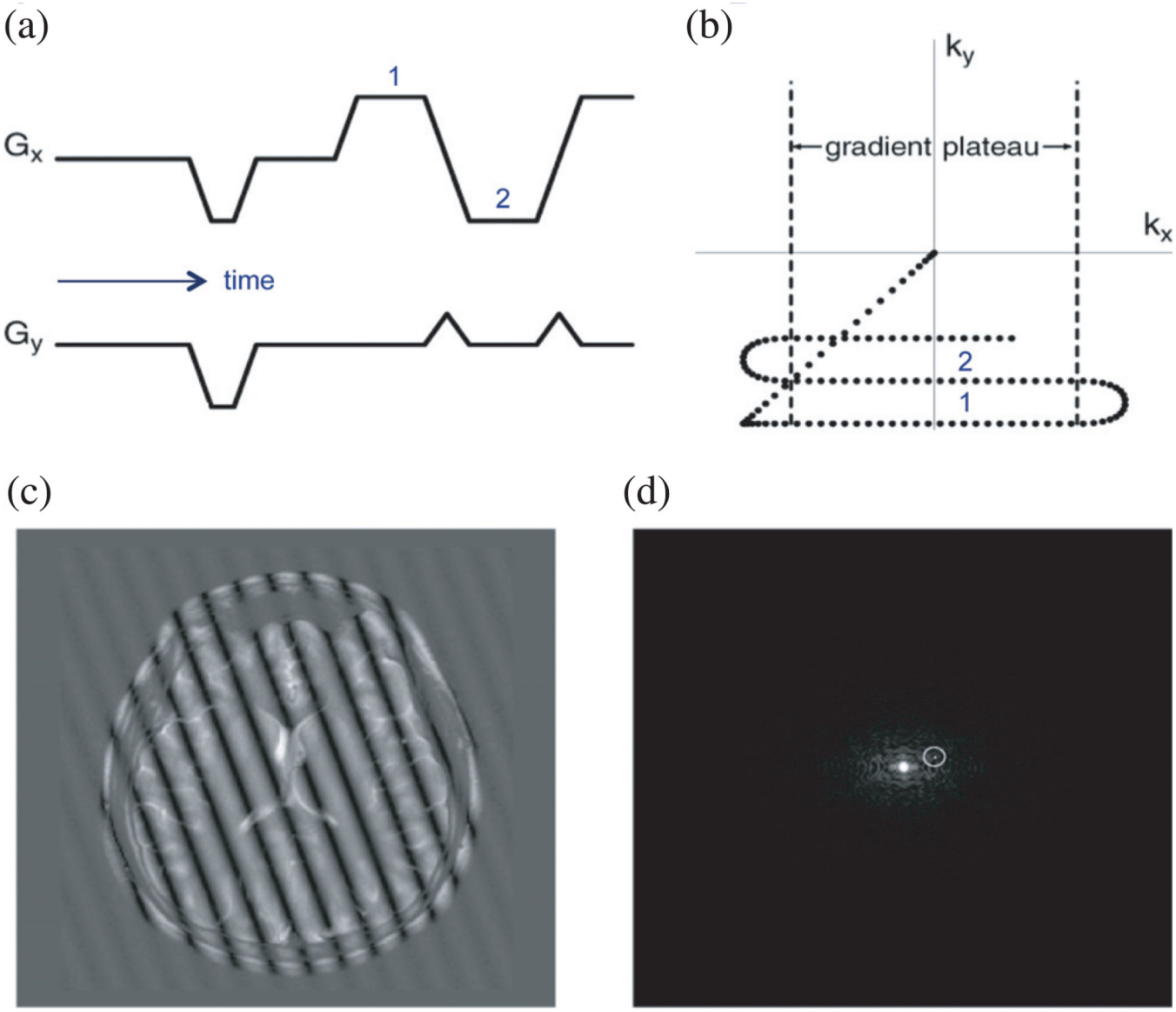

Figure 3.

Basic magnetic field gradient pulses for EPI. Applying a pattern of pulsed linear field gradients along different spatial axes ((*a*) shows gradient amplitude as a function of time) produces sinusoidal spatial modulations of the local signals such that the net signal from the slice traces out a trajectory in the FT space (*k*-space) of the image (*b*). The image reconstruction is then a 2D FT of the acquired data. For example, the spatial contribution to the image (*c*) of a single point in the measured data ((*d*) with the point indicated by the circle) is a single Fourier component (here emphasized by scaling up the value of that point). Current techniques of image acquisition are considerably more sophisticated, but are based on the ideas illustrated here. Adapted from [77] with permission of the author.

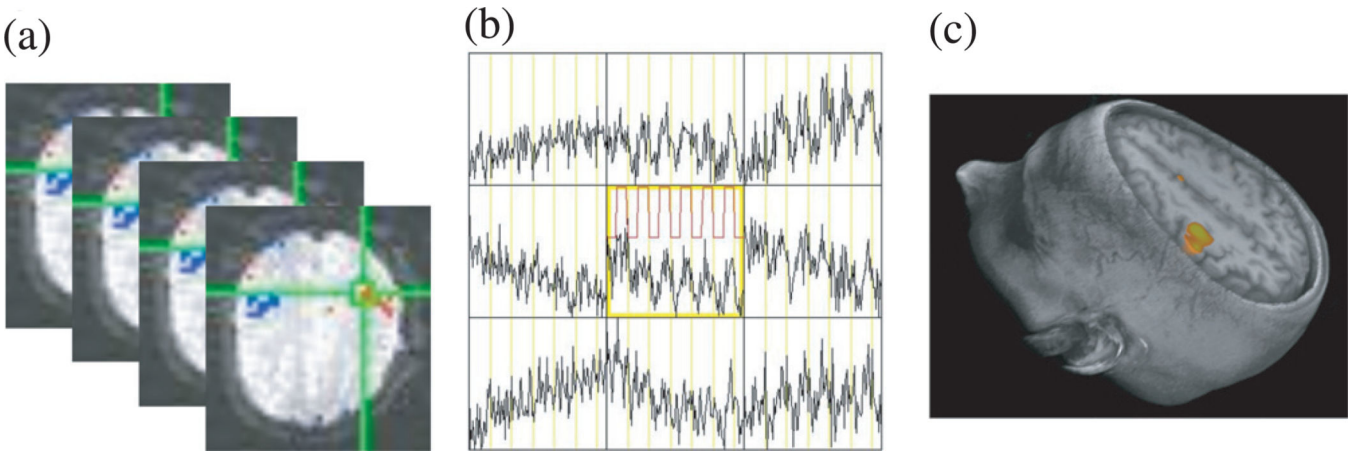

Figure 4.

Basic data acquisition and analysis for fMRI. (*a*) Dynamic T_2^*-weighted images are acquired with a single-shot technique (typically EPI) while a subject performs a task, here illustrated with a simple block design alternating 20 s of finger tapping with 20 s of rest. (*b*) The time course of the signal for each voxel, illustrated here with a 3 × 3 display of the voxels at the intersection of the lines in (*a*), are correlated with the stimulus pattern (shown as the block pattern in the central voxel). (*c*) Voxels with a statistically significant correlation with the stimulus are classed as activated by the stimulus and displayed in color overlay on an anatomical image. As with the image acquisition methods, current fMRI data analysis methods have become considerably more sophisticated, but the idea of correlation as a measure of association remains. In *resting state* methods there is no external stimulus, and instead the fluctuations in the BOLD time course for different voxels are correlated with each other to identify covarying RSNs.

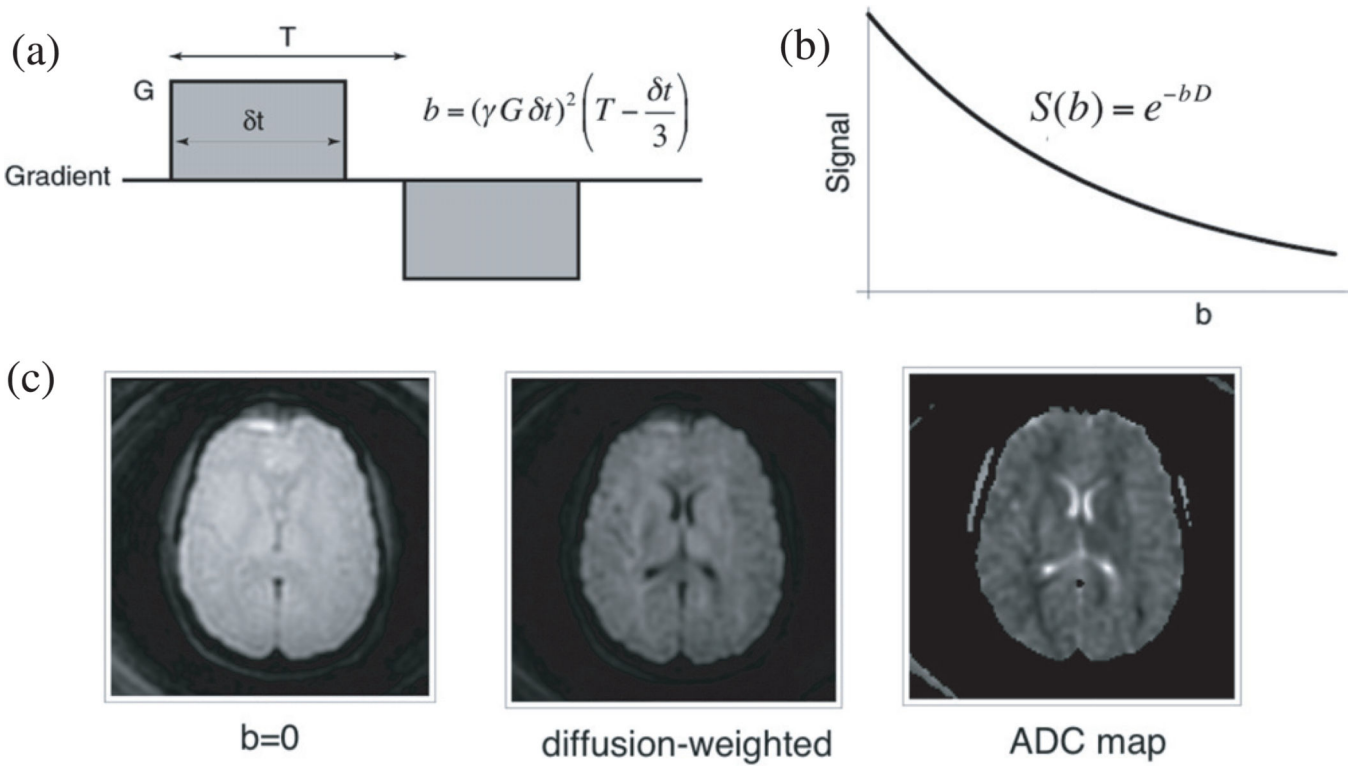

Figure 5.

Diffusion-sensitive imaging. The MR signal can be sensitized to the local random diffusion of water molecules with a bipolar gradient pulse (*a*) that attenuates the measured signal (*b*) by an exponential in *bD*, where *b* depends on the gradient strength and timing parameters, and *D* is the local diffusion constant. This approach is sensitive to displacements of water molecules due to diffusion that are on the order of 10 μm, far smaller than the voxel resolution of the images. (*c*) Images are shown without the bipolar gradient pulse, with diffusion weighting, and the calculated ADC. The direction of the applied gradient pulse is arbitrary, and from measurements of multiple directions the local diffusion pattern can be determined (in the simplest case, the *diffusion tensor*). Diffusion in white matter is highly anisotropic due to the microscopic fiber architecture, and this has led to sophisticated techniques for mapping white matter fiber tracts to provide measures of anatomical connectivity between different brain regions. Adapted from [77] with permission of the author.

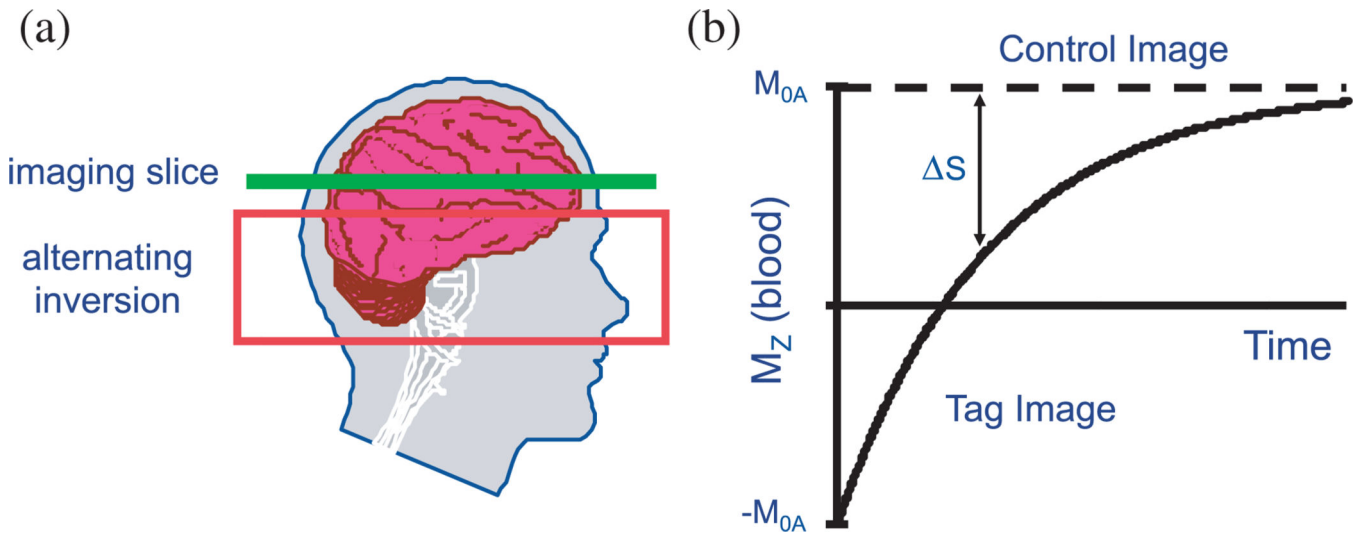

Figure 6.

ASL to measure blood flow. (*a*) Magnetization of arterial blood is alternately manipulated by applying an RF inversion pulse (tag image) or leaving it relaxed (control image). (*b*) After a sufficient delay to allow the labeled blood to be delivered to a slice of interest, the signal difference (control–tag) subtracts out the static signal from the slice leaving a signal proportional to the volume of arterial blood delivered to each voxel during the delay time, providing a quantitative measurement of CBF.

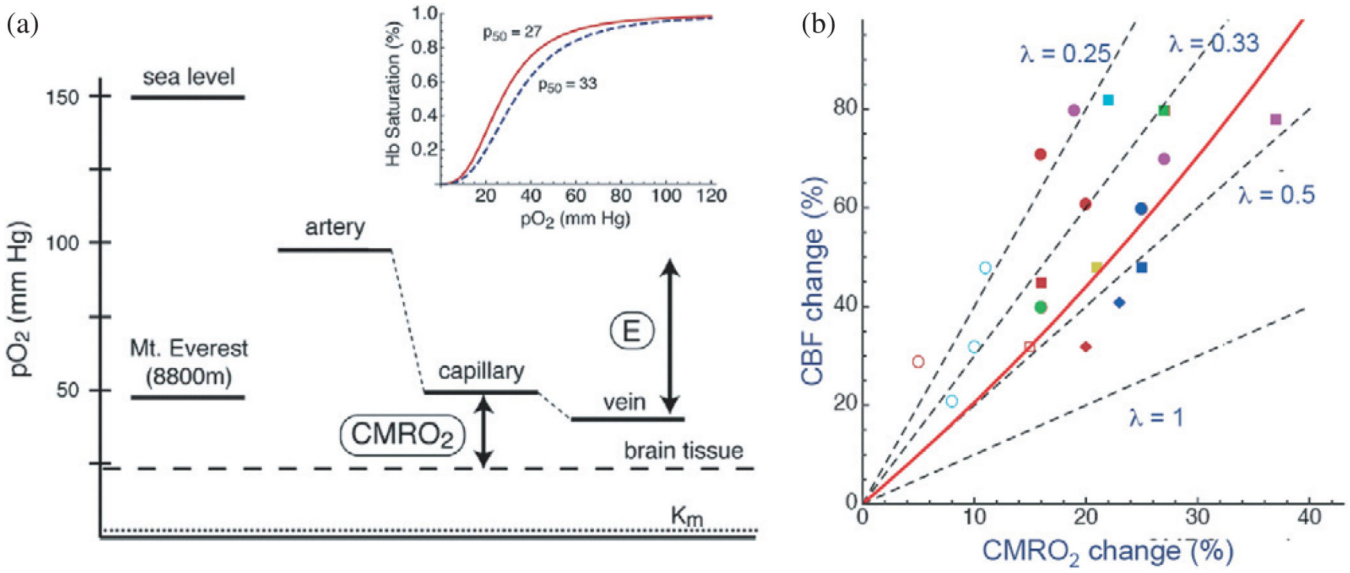

Figure 7.

Possible physiological origin of the mismatch of blood flow and oxygen metabolism changes as a mechanism to maintain tissue pO_2. (*a*) Physical and physiological variations in the partial pressure of oxygen (pO_2), with the O_2 saturation curve of hemoglobin as an inset. To maintain constant tissue pO_2 with increased oxygen metabolism ($CMRO_2$), the capillary pO_2 must increase to increase the diffusion gradient, and this means that the O_2 extraction fraction (*E*) must decrease. The reduction in *E* with brain activation is the origin of the BOLD effect. (Adapted from [77] with permission of the author.) (*b*) Observed fractional changes in blood flow (CBF) and $CMRO_2$ from a number of activation studies, with lines of constant ratio of fractional changes in $CMRO_2$ to CBF (λ, with $\lambda < 1$ indicating a decrease in *E*). The solid line is a modeling prediction of the CBF/$CMRO_2$ coupling ratio needed to preserve tissue pO_2 (adapted from [19]).

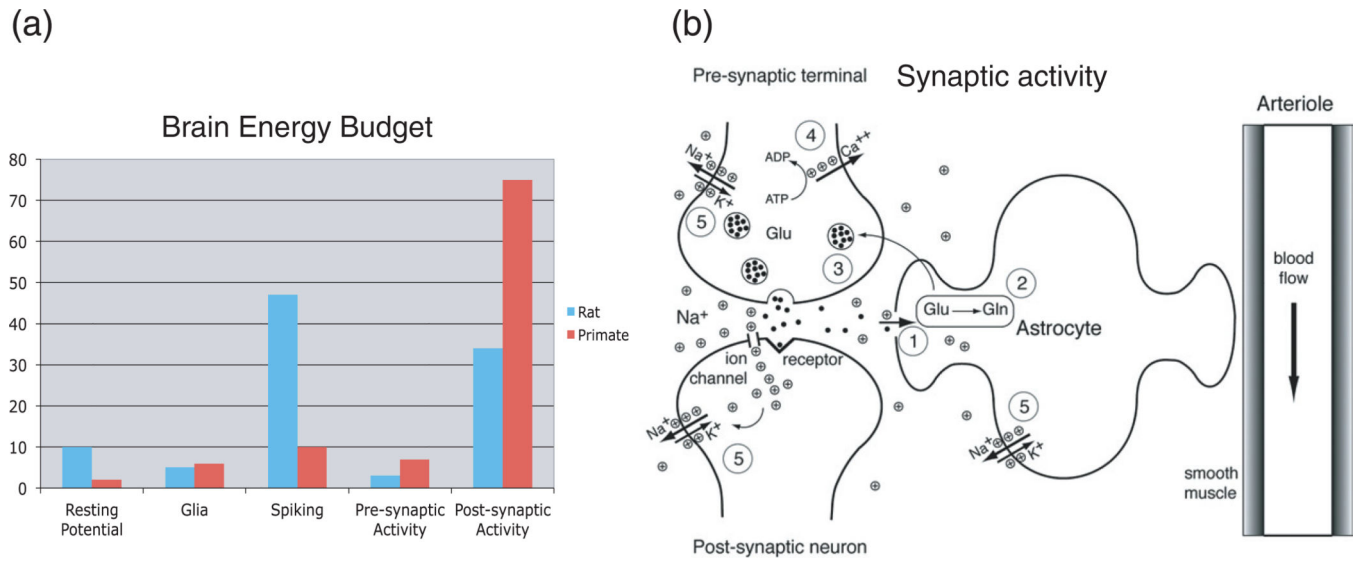

Figure 8.

The energy cost of neural activity. (*a*) Estimates of the ATP consumed for different aspects of neural activity from [80]. Primates have more synapses per neuron, resulting in the dominant energy cost being recovery from synaptic signaling. (*b*) The primary excitatory synaptic signaling involving pre-synaptic Ca^{2+} influx, release of neurotransmitter (glutamate, Glu), opening of post-synaptic Na^{+} channels, and inward Na^{+} currents are all thermodynamically downhill events, and the recovery from these events requires energy metabolism: clearing neurotransmitter through the astrocytes (1), conversion to glutamine (Gln) (2), release of glutamine, uptake by the pre-synaptic terminal, conversion back to glutamate and repackaging the neurotransmitter in vesicles (3), pumping out Ca^{2+} (4) and pumping out Na^{+}. The last event consumes the most ATP, consistent with post-synaptic Na^{+} influx acting like an amplifier of the initial signaling. (Adapted from [77] with permission of the author.)

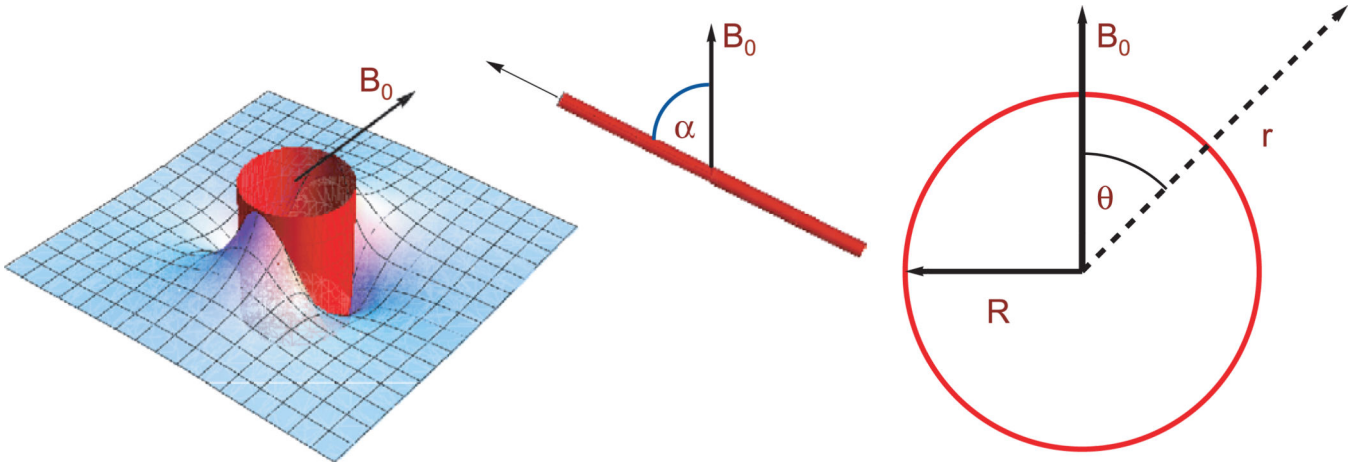

Figure 9.

Magnetic field distortions around a magnetized cylinder. This is the basic physical model for the extravascular effects of a blood vessel containing deoxyhemoglobin, with the dipole pattern of distortions on the left and the geometry of equation (4) illustrated in the other panels.

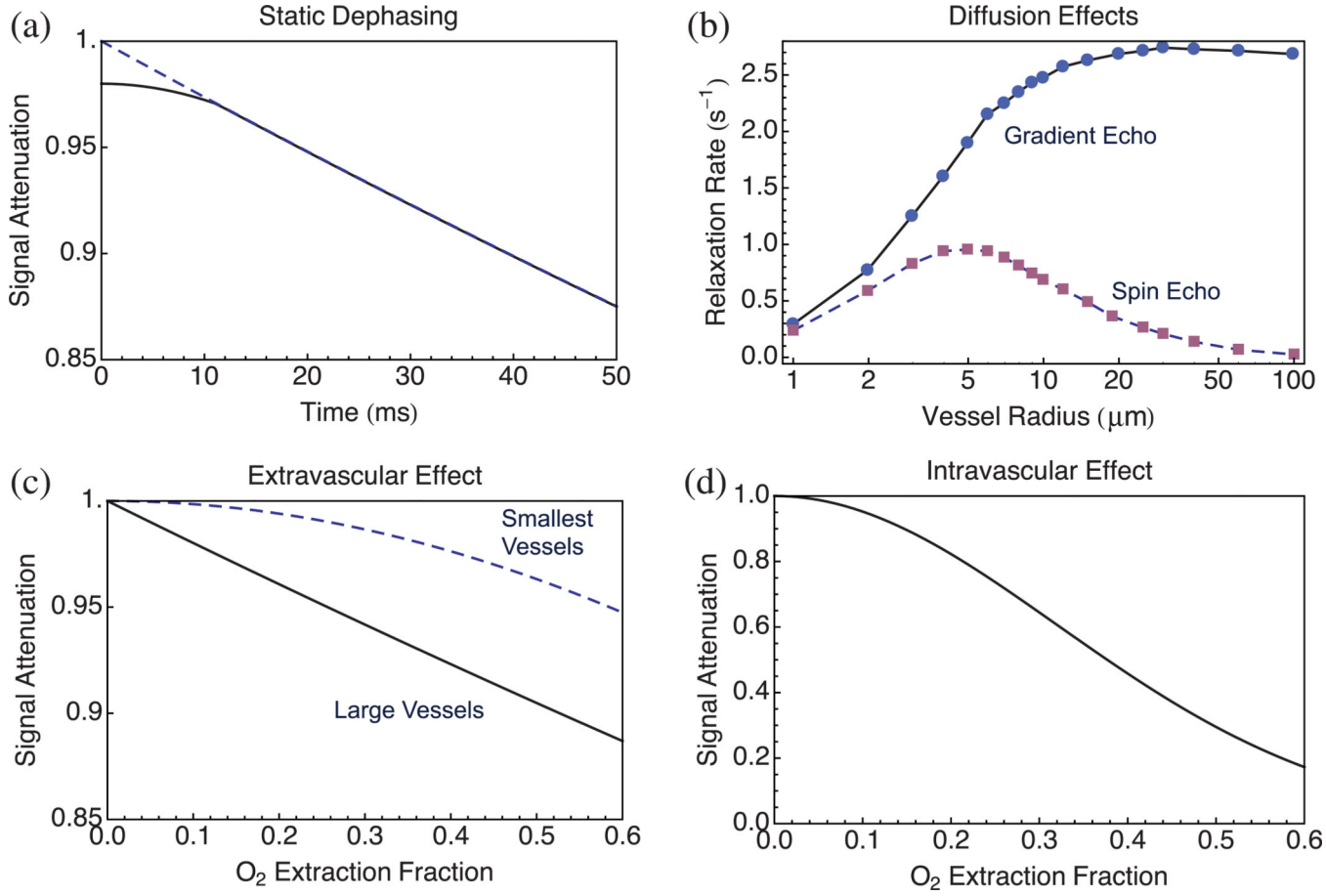

Figure 10.

Model curves of signal decay effects due to magnetized venous blood vessels. (*a*) For static dephasing (no diffusion) the extravascular signal initially decays slowly but then settles to an exponential decay with R_2' given by equation (7). Projecting this portion back to $t = 0$, the difference with the actual curve is the blood volume fraction (0.02 in this example). (*b*) When effects of diffusion are included, the change in relaxation rate depends on the vessel size, with motional averaging reducing the net effect for the smallest vessels (data from Monte Carlo simulations reported in [131] with TE = 30 ms for the GE curve and TE = 60 ms for the SE curve). (*c*) Extravascular signal attenuation as a function of the O_2 extraction fraction for a population of large vessels with radius > 10 μm and a population of the smallest vessels with radius = 2.5 μm, each with the same total volume of deoxyhemoglobin, calculated from equations (7) and (8). (*d*) Intravascular signal as a function of O_2 extraction fraction based on experimental curves measured at a field strength of 3 T [145]. Curves in (*a*), (*c*) and (*d*) were calculated to be consistent with the GE curve in (*b*) with $B_0 = 3$T, TE = 30 ms, OEF = 0.4, $V = 0.02$, and $\delta\omega_0 = 200$ rad s^{-1} ($\nu = 32$ s^{-1}).

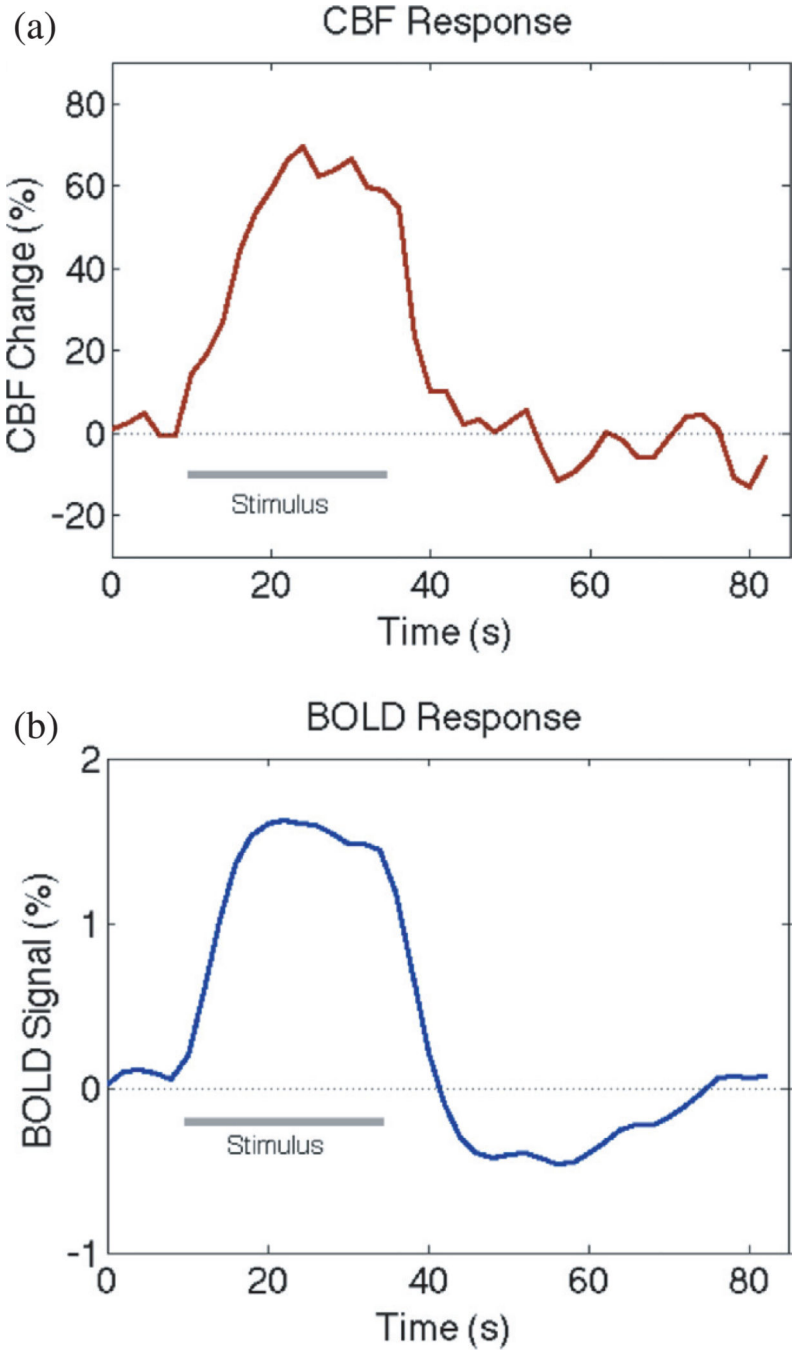

Figure 11.

The BOLD post-stimulus undershoot in human visual cortex. The CBF response (*a*) and the BOLD response (*b*) to a 24s visual stimulus show a strong positive response to the stimulus plus a prominent post-stimulus undershoot of the BOLD signal. Data are from a study comparing luminance and color stimuli which found no difference in the responses [229], and the data for the two types of stimulus were combined for these curves. The origin of the

BOLD undershoot is still debated, and it could potentially be due to vascular or metabolic effects.

Table 1

Physiological variables related to neural activation. The standard abbreviations used in physiology, and the symbols used in the equations in this paper are given in the first column. Typical baseline values are for healthy human adults, but may vary significantly. The fractional changes with activation vary widely with different stimuli; the values given are meant to suggest typical relative values and not standard absolute values. The units chosen are convenient for the later modeling discussion, but not necessarily the standard units used in the physiology literature.

Variable	Description	Typical baseline value	Example change with activation
CBF (F)	Cerebral blood flow (ml arterial blood per ml tissue per min)	0.5 min^{-1}	0.65 min^{-1} (+30%)
OEF (E)	Oxygen extraction fraction (dimensionless)	0.4	0.34 (−15%)
$CMRO_2$ (R)	Cerebral metabolic rate of oxygen (micromoles O_2 per ml tissue per min, or mM min^{-1})	1.6mM min^{-1}	1.8mM min^{-1} (+12%)
CMRGlc	Cerebral metabolic rate of glucose (micromoles glucose per ml tissue per min, or mM min^{-1})	0.3mM min^{-1}	0.4mM min^{-1} (+30%)
CBV	Cerebral blood volume (dimensionless fraction of tissue volume)	0.05	0.055 (+10%)
$[O_2]_a$	Arterial oxygen concentration (micromoles O_2 per ml blood)	8mM	8mM (—)