Simultaneous EEG and fMRI
Recording, Analysis, and Application

EDITED BY:

Markus Ullsperger, MD, PhD
Biological Psychology / Donders Centre for Cognition
Radboud University Nijmegen
Nijmegen, The Netherlands; and
Max Planck Institute for Neurological Research
Cologne, Germany

Stefan Debener, PhD
Institute of Psychology
Carl von Ossietzky University Oldenburg
Oldenburg, Germany; and
Biomagnetic centre
Hans Berger Clinic for Neurology
University Hospital Jena
Jena, Germany

OXFORD UNIVERSITY PRESS
2010
Introduction

fMRI and other non-invasive imaging methods have greatly expanded our knowledge of human brain function. Although MRI was invented in the early 1970s (Lauterbur, 1973) and has been used clinically since the mid–1980s, its use in cognitive neuroscience expanded greatly with the advent of blood oxygenation level dependent (BOLD) functional imaging, and by now fMRI is a mainstay of neuroscience research. The BOLD contrast is based on an increase in blood oxygenation and flow to activated regions. In the late 1800s Mosso first observed that neural activation increases blood flow to the activated brain region (see Iadecola, 2004, Zago et al., 2009, or Raichle, 2000, for review). Ogawa et al. (1990) and Belliveau et al. (1990) first used MRI to measure such changes in cerebral flow and oxygenation in rats. Belliveau et al. (1990) used an exogenous contrast agent to measure changes in blood volume, but Ogawa et al. (1990) showed that deoxyhemoglobin acts as an endogenous contrast agent. By manipulating O₂ and CO₂ levels, they showed that excess deoxygenated blood leads to signal loss near vessels, which could be reversed by increasing the flow of freshly oxygenated blood. Soon thereafter it was shown in humans that changes in normal brain function lead to changes in blood oxygenation and flow that can be measured with MRI (Belliveau et al., 1991; Kwong et al., 1992; Ogawa et al., 1992). The origin of the functional BOLD contrast is actually an oversupply of freshly oxygenated blood that leads to a signal increase in the activated areas. Compared to other functional imaging methods, fMRI offers much higher spatial resolution and it is entirely non-invasive, which has allowed us to improve the mapping of the functional parcel- lation in human brain.

fMRI activation is usually without much question interpreted as a marker of neural function. But as the fMRI signal is a hemodynamic signal, it is only an indirect marker, and our ability to interpret fMRI in terms of neural function depends on how well the fMRI signal reflects the neural activation. For example, it is not known if the different neural processes are all represented equally in the fMRI signal, or whether some are overrepresented or underrepresented. This is still relatively unknown territory that many different groups are currently trying to address with multidisciplinary approaches.

There has been over 50 years of neurophysiological work in nonhuman primates and other animals, often painstakingly mapping the functional properties of isolated neurons. And although this has offered a wealth of knowledge about brain function, for obvious reasons a direct extrapolation to human brain function is difficult. In recent years there has been an increasing amount of work aimed at elucidating how the neural signals that are measured invasively with intracortical micro-electrodes in animals relate to the non-invasive measures of brain function that can be researched in humans, like fMRI, electroencephalography (EEG) and magnetoencephalography (MEG). The goal would be to arrive at measures for non-invasively interpreting neural functioning at high resolution in humans. Clearly, fMRI is an excellent candidate for such an endeavor, but because fMRI measures blood-flow, it can only be used as a surrogate measure of neural activation. To allow us to better relate the body of fMRI work in humans to the bulk of neurophysiological work done in animals, it is necessary to determine how fMRI signals relate to neural activity. This can be done by combining invasive methods with fMRI in monkeys or other animals, either simultaneously (Goense and Logothetis, 2008; Logothetis et al., 2001) or consecutively (Maandag et al., 2007; Disbrow et al., 2000; Lipton et al., 2006; Tsao et al., 2006; Smith et al., 2002; Lu et al., 2007). Neurophysiological recording in humans also offers opportunities to study directly the relationship between neural signals and fMRI (Privman et al., 2007; Mukamel et al., 2005).
the temporal lobe, and is usually conducted in a surgical or presurgical context (Engel et al., 2005; Privman et al., 2007; Mukamel et al., 2005). Another approach to a better understanding and interpretation of the human fMRI studies is to perform comparative fMRI in humans and monkeys, which allows us to look at homology questions and provides a better basis for extrapolating animal work to human work (Nakahara et al., 2002; Koyama et al., 2004; VanDuffel et al., 2002; Tsao et al., 2003; Sawamura et al., 2005; Rajimhreh et al., 2009; Tsao et al., 2008; Denys et al., 2004).

As both electrophysiological and imaging methods have their strengths and limitations and measure only particular aspects of brain function, the integrative approach allows us to obtain a more complete picture. This is expected to improve our interpretation of both methods. The most obvious limitation of intracortical microelectrodes is its highly restricted spatial sampling; for EEG it is the limited spatial resolution of its reconstructed sources (which is an ill-posed inverse problem; cf. Chapters 1.1 and 3.8) and for fMRI its limited temporal resolution, which also stems from the fact that it is a hemodynamic signal and does not measure the neural events themselves.

The goal of this chapter is to give an overview of the relation between the BOLD signal and the underlying neural signals. Our focus is mainly on intracortically recorded neural signals, recorded with microelectrodes. Neural signals that are recorded are spikes, or single- and multi-unit activity (SUA and MUA) and local field potentials (LFP). The increases in local neural activity upon stimulus presentation, and the concomitant increased energy demands of neurotransmission and spiking, lead to an increase in blood flow to the activated area, which ultimately drives the BOLD response. Logothetis et al. (2001) showed that the BOLD signal is not as well correlated to single-unit activity, but correlates better with the LFP, a mesoscopic signal that includes membrane potential fluctuations, oscillations, and postsynaptic and presynaptic events. This raises questions about whether all aspects of neural activity drive the BOLD response equally, and if not, which ones are more important—for instance, the input versus the output from an area, or inhibition versus excitation, or stimulus-driven or neuromodulatory activity. Answers to such questions will directly affect our interpretation of fMRI results and help us to better understand results obtained with fMRI. We will discuss which neural events are thought to drive the hemodynamic response, but to get some insight in the coupling between neural activity and BOLD, we will also discuss neurovascular coupling and the specificity of the BOLD response, as these issues have direct bearing on our understanding of the coupling between BOLD and neurophysiological signals.

**BOLD Contrast Mechanism**

Ogawa et al. (1990) first discovered the BOLD contrast in rats, when they observed that the intensity of the vascular signal in gradient-echo (GE) images decreased when blood was deoxygenated, and increased when the flow of freshly oxygenated blood increased. Soon after, it was shown that that the BOLD contrast could be used to detect functional activation in humans (Kwong et al., 1992; Ogawa et al., 1992). When neural activity in a given brain area increases, via stimulation or task performance, this triggers an increase in the flow of fresh blood to the activated area in order to meet the increased metabolic demands. The BOLD contrast is based on the concentration of deoxyhemoglobin in the blood, since deoxyhemoglobin is paramagnetic and acts as a contrast agent: an increase in its concentration decreases the relaxation time ($T_2^*$). The increase in oxygen supply to the active tissue is more than the oxygen that is used by the neurons, and hence there is a relative increase in the oxyhemoglobin concentration, and a decrease in the deoxyhemoglobin concentration. This increases $T_2^*$ and leads to a signal increase in the GE images (Kwong et al., 1992; Ogawa et al., 1992). Belliveau et al. (1990; 1991) showed functional MRI based on similar principles but using an exogenous contrast agent that is sensitive to changes in blood volume.

Although the first fMRI experiments were performed in early 1990s, what exactly the BOLD signal represents is still unclear. Questions remain not only about which neural or metabolic changes exactly trigger the BOLD signal, but also about the relative contributions of flow increases versus oxygen extraction, or which parts of the vascular tree contribute most to the BOLD signal. And given these uncertainties, the BOLD signal is not (yet) suitable as a quantitative metric for brain function.

**Properties of the BOLD Signal**

The presence of paramagnetic deoxyhemoglobin (dHb) in the blood leads to susceptibility gradients, which are local variations in the magnetic field. These gradients exist near vessels because of their high dHb content, and their size depends on the vessel size and dHb concentration among other things. Spins within these gradients experience dephasing, which leads to $T_2^*$-based signal loss. Based on theory and simulations (Boxerman et al., 1995; Kennan et al., 1994; Weisskoff et al., 1994) it has been determined that with the typically used GE-based sequences, one observes BOLD signal arising from different vessel sizes ranging from capillaries to large veins. Large draining veins can have very strong BOLD signals, but they are downstream from the neural activation and can be quite remote from the activated area. However, the relative contribution of large and small vessels to the BOLD signal is still debated. One of the reasons why there is still debate about this is that the properties of the BOLD signal depend strongly on MR-hardware (most obviously field strength) and acquisition parameters (Duong et al., 2003; Goense and Logothetis, 2006; Jin et al., 2006; Yacoub et al., 2001), which complicates the comparison of results across labs or studies. For example, it is well established that at low magnetic field (e.g., 1.5T) the BOLD signal is very sensitive to large vessels. This sensitivity...
progressively decreases at higher field (Yacoub et al., 2001). But although the vessel-fraction is decreased at high field, high-resolution studies show that also at high field there can still be a substantial large vessel contribution (Harel et al., 2006; Jin et al., 2006; Zhao et al., 2004). Because large-vessel BOLD signals decrease the accuracy with which functional activity can be localized, it is important to decrease the sensitivity to large-vessel signal as much as possible.

Figure 1.2.1 shows the BOLD signal elicited by visual stimulation in monkey visual cortex acquired at 7T. Functional activation is seen in the entire early visual cortex (V1–V5) and at this resolution it can easily be seen that the BOLD signal occurs at both the cortical surface and in gray matter, with the strongest BOLD signals at the cortical surface and in the calcarine sulcus. This is the large-vessel contribution, which can be more easily distinguished at high resolution. Note that often statistical maps (instead of percentage change maps) are used to display activation, and they depict correlation with the stimulus. These do not necessarily show the highest p-values at the surface, because although vessels typically have larger signals they also tend to have higher noise (Goense and Logothetis, 2006). The sensitivity to large vessels has the drawback that an area of activation

\[\text{Figure 1.2.1. Functional activation acquired in visual cortex of an anesthetized monkey at 7T at a spatial resolution of 500x500x2000 µm}^3\]. The map shows the percent signal change in response to a full-field (30°) black-and-white rotating checkerboard presented to both eyes. Functional activation is seen in visual areas V1–V5. The scanner and anesthesia procedures have been described previously (Logothetis et al., 1999; Pfeuffer et al., 2004), volume transmit coil, 30 mm receive coil, 8-segment GE-EPI, TE 20 ms, TR 750 ms.
identified in a functional map could in fact be a vessel remote from the area of neural activity. Also, when a blood vessel in a sulcus shows activation, at low resolution it could lead to ambiguity about the exact location of the activated tissue.

Many details of the BOLD signal are still poorly understood, in particular the relative contributions of venous, arterial, and capillary fractions to the BOLD signal, and the relative contributions of blood flow increases and oxygen consumption (Haacke et al., 2001; Buxton et al., 2004; Uludag et al., 2009). This is the case for the positive BOLD response, for the initial dip (the existence of which is still debated), the poststimulus undershoot, as well as for fMRI signals recorded with different methods (Zhao et al., 2007; Yacoub et al., 2006; Ugurbil et al., 2006; Kim et al., 2007). A discussion of the different fMRI methods and their properties is beyond the scope of this chapter. Here we note only that the general consensus is that GE-BOLD represents mostly a venous signal, which becomes more strongly weighted toward smaller venules and capillaries as the field strength increases, spin-echo (SE) BOLD represents mostly a capillary signal, the cerebral blood volume (CBV) signal is thought to represent smaller vessels (arteries and veins) and capillaries, and the cerebral blood flow (CBF) signal is thought to represent mostly arterioles and capillaries.

Spatial Resolution and Specificity of fMRI

How well the functional activation is localized to the actual place of neural activation depends on the achievable fMRI resolution and the specificity of the hemodynamic signal. The achievable spatial resolution is determined by scanner hardware and the signal-to-noise ratio (SNR). The specificity is determined by the fMRI method that is used and how closely the hemodynamic response reflects the actual neural activity. BOLD signal originating from large vessels may be remote from the site of activation, but BOLD signal from capillaries can reasonably be assumed to be closely related to the neural activity in that area. Hence, the specificity is not only determined by biological factors, but also by the choice of fMRI method, hardware, and sequence parameters (Harel et al., 2006b; Ugurbil et al., 2003).

With advanced scanner hardware the spatial resolution achievable for structural imaging in vivo is of the order of 200–300 μm in-plane for whole-head imaging in humans (Wald et al., 2006; Ugurbil et al., 2006) and ∼100 μm for localized imaging (Nakada et al., 2005). In animals, resolutions of 70–100 μm have been achieved in macaques (Logothetis et al., 2002) (Figure 1.2.2), a few tens of μm in rodents, and even higher resolutions in vitro (Ciobanu and Pennington, 2004; Fu et al., 2005). In principle fMRI can also be done at such resolutions, although it is constrained by the relatively low contrast-to-noise ratio (CNR) of the functional activation and the limited amount of time available for the acquisition of each image. Because of its speed, echo-planar imaging (EPI) (Mansfield, 1977) is typically used for fMRI (it can collect one image per excitation or repetition time TR). However, EPI requires high-performance hardware, and the limitations of the gradients often restrict the maximally achievable resolution. Despite this, by using segmented EPI

Figure 1.2.2. Anatomical image of macaque V1. The high-resolution GE image at a resolution of 100x100x1000 μm³ and volume of 0.01 μl shows the small perpendicular intracortical veins and layer IV, which is indicated by the Gennari line. The Gennari line has a higher myelin content than the rest of the cortex. Technical data: vertical 4.7T scanner, described in Logothetis et al. (1999), volume transmit coil with 4-channel receive array, matrix 768x512x22, TE 23.5 ms, TR 2000 ms.
or parallel imaging, high resolution fMRI can be done, and in monkeys functional maps at 125 μm in-plane resolution have been shown (Logothetis et al., 2002), while in rats, maps with 50–100 μm have been demonstrated (Silva and Koretsky, 2002; Xu et al., 2003). In human fMRI studies, typical resolution is ~3x3x3 mm³, and currently the highest resolution achieved is about 500 μm in-plane (Pfeuffer et al., 2002; Yacoub et al., 2005). The magnitude of the functional activation in gray-matter is only a few percent, and although image SNR decreases for smaller voxels, the functional signal tends to actually increase as the voxel size decreases, due to a decrease in partial volume effects. At a few hundred μm or less, the fMRI resolution can be higher than the point spread function (PSF) of the activation. Hence, the theoretically achievable spatial resolution is probably limited by the spatial extent of the neural signals and the hemodynamic regulation.

At high resolution specificity is important to be able to visualize structures like patches, columns, or layers. But also at low resolution increasing the specificity of the functional signal is important to eliminate the effect of draining veins and thus to increase the accuracy of the mapping. Specificity depends on (1) the anatomy of the capillary bed, (2) which fMRI method is used, and (3) neurovascular coupling and the spatial scale of the regulation of blood flow.

**Anatomy of the Cortical Vascular System**

Since the fMRI signal is a blood-flow-dependent signal, its specificity ultimately depends on the anatomy of the vascular bed and the spatial scale of blood-flow regulation. Figure 1.2.3 shows a corrosion cast of the cortical vasculature in macaque V1. The cortical blood supply is characterized by large arteries that run along the surface of the cortex, branch into smaller pial arteries, and branch finally into intracortical arterioles (100–200 μm) that enter the cortical gray matter perpendicular to the surface. These intracortical vessels eventually branch into capillaries, and blood is collected in intracortical venules of 100–200 μm, which form larger venules and veins on the surface of the cortex (Duvernoy et al., 1981). The distance between capillaries at any place in the gray matter of the cortex is about 40 μm (Weber et al., 2008). Gray matter is more highly vascularized than white matter, and there are differences in vascular density between different cortical areas; areas with higher vascularization, like primary sensory cortex, typically also show higher fMRI responses.

The differences in cerebral blood volume in vivo are shown in Figure 1.2.4. The steady-state CBV image shows the change in relaxivity $\Delta R_2^*$ induced upon injection of the intravascular iron-based contrast agent MION (monocystalline iron oxide nanocolloid), which is a function of blood

---

**Figure 1.2.3.** Corrosion cast of monkey V1 showing the microvasculature of the cortex (from Weber et al. [2008], with permission). Arteries and large draining veins are located in the pial layer. Penetrating arterioles and venules can be seen in addition to the capillary network. Intracortical arteries are shown in red and intracortical veins in blue. The vascular density in white matter is lower than in the gray matter.
volume (Dennie et al., 1998; Wu et al., 2004). The figure shows the higher blood volume in gray matter than in white matter as well as the high blood volume at the surface of the cortex. But there are also differences in capillary density within the cortex, for instance laminar differences (Lauwers et al., 2008; Weber et al., 2008), which are reflected in differences in perfusion. Figure 1.2.4 shows these differences in macaque V1, indicating that the middle cortical layers have higher blood volume (Goense et al., 2007).

Regulation of Cortical Blood Flow

Because the brain has high energy demands, it needs a relatively constant supply of nutrients and oxygen and is sensitive to blood-flow changes. To protect the brain from injury, cerebral blood flow is tightly regulated at multiple levels, from macroscopic to microscopic (Faraci and Heistad, 1998; Hamel, 2006; Iadecola, 2004). An important aspect of cerebrovascular regulation is its autoregulation, whereby cerebral blood flow is kept within the normal range despite changes in systemic blood pressure.

Cerebral blood flow is coupled to neural metabolism, meaning that changes in neural activity produce concomitant changes in blood flow. There are multiple mechanisms whereby changes in neural activity and metabolism lead to a change in blood flow. Events triggering changes in perfusion can be the partial pressure of CO₂ and oxygen, the pH, or tissue concentrations of metabolites (Faraci and Heistad, 1998). There are also multiple signaling molecules and pathways involved in the neurovascular coupling, for instance, NO, prostaglandins, etc. (Hamel, 2006; Iadecola, 2004).

The specificity of the BOLD response to localized neural activation depends on the spatial scale of the blood flow regulation. Changes in blood flow are mediated by dilation and constriction of arteries and arterioles, while venules and veins have no smooth muscle and are mostly kept open in the brain. The neurovascular response can be quite localized as evidenced by optical imaging and two-photon microscopy experiments. For instance, ocular dominance columns and blobs in macaque V1, which are of the order of a few 100 μm, have been shown with optical imaging (Ts’o et al., 1990). Chaingneau et al. (2003) showed that blood flow in the rat olfactory system in vivo is regulated at the level of individual glomeruli, and with two-photon microscopy it has been shown that blood flow is regulated at the level of individual arterioles or even capillaries. Neurons and associated astrocytes form the so-called neurovascular units, and astrocytes are thought to play an important role in mediating the blood-flow response (Schummers et al., 2008). Constriction and dilation of individual arterioles were observed to be mediated by astrocytes by inducing local Ca²⁺ increases (Metea and Newman, 2006; Mulligan and Mavricar, 2004; Zonta et al., 2003; Takano et al., 2006). Dilation and constriction of capillaries by specialized structures called pericytes have been shown in the retina (Peppiatt and Attwell, 2004) and pericytes may perform similar functions in the brain, although this is still unclear (Hirase et al., 2004).

Specificity of Different fMRI Methods

Based on the above, the expectation is that the scale of the hemodynamic regulation is currently not the limiting factor for most fMRI applications except possibly for ultra-high resolution fMRI. The specificity of the fMRI signal further depends on the fMRI method that is used, scanner hardware (for instance field strength), and acquisition parameters. Different fMRI methods have different specificity (Weisskoff et al., 1994; Boxerman et al., 1995;
Kennan et al., 1994). We already mentioned that for instance GE-BOLD although sensitive to both vessels and capillaries, at low field is dominated by large vessels, and the contribution of the capillary signal increases at high field. SE-BOLD, CBV, and CBF methods are less sensitive to large vessels (Harel et al., 2006b). The GE signal is sensitive to the phase dispersal near large vessels, i.e., water-protons in voxels near large vessels exhibit a range of phases due to the susceptibility gradients, and the phase dispersal in a voxel causes signal within a voxel to cancel out (called static dephasing). SE is not sensitive to static dephasing because the accumulated phase dispersal is refocused by the 180 degree pulse. However, near capillaries a dynamic effect is dominant which gives rise to the SE-BOLD signal, that is, spins that move within the field gradients accrue

Figure 1.2.5. Specificity of GE- and SE-BOLD fMRI acquired at 4.7T. The GE-BOLD map (A) at a resolution of 333x333x2000 μm³ shows highest activity at the cortical surface and near vessels in the sulcus (Goense et al., 2007; Logothetis, 2008), whereas the SE-BOLD map (B) at a resolution of 250x175x2000 μm³ shows a BOLD signal that is better confined to the gray matter. Laminar structure can be seen in the SE-fMRI signal. Partial volume effects in the slice direction were not detrimental due to the anatomy of monkey V1 and a slice angle that was perpendicular to the cortical surface.
phase dispersal, and the change in phase due to this movement cannot be refocused. This is called dynamic averaging, characterized by the “apparent $T_2$-eff” or $T'_2$ (Ugurbil et al., 2000). In addition, there is a $T_2$-effect that arises from intravascular protons. This is at fields up to 3T the dominant contributor to the SE-BOLD signal (Duong et al., 2003). At higher field, $T_2$ of blood is short and the $T_2$-effect becomes less dominant.

Differences in specificity between the different methods can most easily be seen in high resolution fMRI. Figure 1.2.3 shows that large vessels are only located on the surface of the cortex and that vessels within gray matter are up to ~200 µm in monkeys. In high-resolution GE-EPI fMRI, the largest functional changes occur at the surface and near large vessels (Figure 1.2.5). In contrast, the largest SE-BOLD functional changes occur within the gray matter, approximately in layer IV, and not much functional activation is seen near large vessels, illustrating the improved specificity of SE-BOLD over GE-BOLD. However, if the spatial resolution is 1 mm or lower, signals cannot be clearly differentiated as originating from the surface or from within gray matter because of partial volume effects.

The specificity of the different fMRI methods can also be demonstrated by visualizing cortical columns. Ocular dominance columns (ODCs) in humans (~1 mm in diameter; Adams et al., 2007) are often used, or orientation columns in cats (also ~1 mm in diameter; Lowel et al., 1987). Columnar resolution was successfully demonstrated in humans and cats using GE-EPI (Moon et al., 2007; Cheng et al., 2001; Menon and Goodyear, 1999; Dechent and Frahm, 2000; Yacoub et al., 2008). When GE-BOLD is used, typically subtraction paradigms are used that subtract out the signals common to both stimuli, and hence it removes the nonspecific vessel signals, or alternatively vessel signals are thresholded to ameliorate the predominance of vessel signal in GE-BOLD (Moon et al., 2007; Cheng et al., 2001; Logothetis et al., 2002). The drawback of subtraction paradigms is that orthogonal stimuli are needed, which for many stimuli are not available. With SE-EPI, ocular dominance columns have been shown in humans (Yacoub et al., 2007; Yacoub et al., 2008), and using CBV and CBF methods, single-condition maps of orientation columns were demonstrated in cat V1 (Duong et al., 2001; Zhao et al., 2005). These more specific methods also allow functional mapping of laminar differences within the cortex, as demonstrated in V1 of cats and monkeys (Goense et al., 2007; Harel et al., 2006a; Zhao et al., 2006; Zappe et al., 2008).

**BOLD Temporal Resolution**

The temporal resolution of the BOLD signal is determined by the hemodynamic response, which is relatively slow, and hence fMRI cannot capture quick neuronal changes. The hemodynamic response takes a few seconds to develop and reaches its peak between 5 and 10 s after stimulus onset. Unless special paradigms are used, for instance, making use of the prolonged hemodynamic response to brief and strong stimuli (Ogawa et al., 2000) the useful temporal resolution of fMRI is in the second-range.

**The Neural Signals**

To be able to answer the question how BOLD relates to the neural signals and what aspects of the neural signals are best represented by the fMRI signals, we need to take a closer look at what we actually mean when we talk about “the neural signals.” Usually the assumption is that they represent signals measured by intracortical electrodes or microelectrodes with standard extracellular methods. And we do wish to compare the BOLD signal to such electrophysiological recordings because a large amount of our knowledge about neural and brain function is based on these extracellular techniques.

Typically single- or multiple-unit activity is recorded with microelectrodes, and we look at the specificity and spatial extent of these signals, for example, what types of neurons are recorded and how large is the area that is sampled. However, what is recorded by an electrode is not always fully representative of all the processes that occur in the brain because electrodes measure only a subset of the neural processes.

The signal measured by an electrode placed at a neural site is the mean extracellular field potential (mEFP) from the weighted sum of the electrical sinks and sources along multiple cells. Its waveform is characterized by fast action potentials superimposed on relatively slowly varying field potentials. Different signals are recorded depending on the impedance of the electrode. If a microelectrode with a small tip is placed close to the soma or axon of a neuron, then the measured mEFP directly reports the spiking of that neuron and frequently also that of its immediate neighbors. With electrodes of the order of several 100s of kΩ, single spikes or single-unit activity (SUA), multi-unit activity (MUA), and local field potentials (LFP) are recorded. Low impedance electrodes record predominantly LFP, while sharp electrodes/pipettes are typically used to record single units, and not much LFP is observed.

**Single- and Multi-Unit Activity**

The standard technique in neurophysiological research has been the recording of single spikes. Single spike monitoring has the best possible spatial and temporal resolution and it has been and it will continue to be the method of choice when single cell properties are the subject of investigation. It provides information of the spike output of the isolated cell and its response properties, for instance its receptive field or its tuning to different stimuli. Depending on the location and impedance of the electrode, often multiple neurons are recorded simultaneously. If MUA is recorded, the spikes generated by different neurons can be sorted based on their...
shape. The spike shape that is recorded can vary depending on the location of the electrode with respect to the neuron. However, for accurate sorting, tetrodes (4-contact electrodes) or multicontact electrodes are advantageous or often necessary, particularly when the spike shapes of the neurons are similar.

A drawback of single- and multi-unit recordings is that they suffer from an element of bias toward certain cell types and sizes (Stone, 1973; Towe and Harding, 1970). Figure 1.2.6 shows a drawing of the cortical circuit and illustrates the variation in size and morphology of neurons within the cortical sheet. The measured spikes however mostly

Figure 1.2.6. Schematic of the modular arrangement of cortical cells and connections (reproduced from Szentagothai [1978], with permission). The cortex has a columnar and well-ordered parallel structure. The drawing shows the different types of neurons and connections. Pyramidal cells, spiny stellate (sp.st.), small and large basket cells (s.b.c. and l.b.c.), microgliform cells (m.g.), chandelier cells (ch.c.), Martinotti cells.
represent only very small neural populations of large cells, which are by and large the pyramidal cells in cerebral cortex and Purkinje neurons in cerebellar cortex. The magnitude of EFPs in the MUA range, for example, was shown to be a function of cell- and axon size (Gur et al., 1999). Combined physiology-histology experiments also demonstrated that the magnitude of MUA is site- (Buchwald and Grover, 1970) and cell-size specific (Nelson, 1966), varying considerably from one brain region to another (e.g., neocortex vs. hippocampus) but remaining relatively constant within a particular region. Homogeneous populations of large cells were found to systematically occur at sites of large-amplitude fast activity and vice versa (Grover and Buchwald, 1970). Similarly, the magnitude of axonal spikes is directly correlated with the size of the transmitting axon (Hunt, 1951; Gasser and Grundfest, 1939).

Recording from nonpyramidal cell types, for instance interneurons, is often difficult both because of their small size and because their response is often uncorrelated to the stimulus or to the behavioral state of the animal. Since response to a stimulus is often the criterion for successful isolation of a neuron or cluster of neurons, this can lead to a bias against certain neurons, for instance inhibitory neurons, against neuromodulatory neurons, or against neurons that have very low firing rates. This could introduce a substantial bias because there is reason to believe that neurons with very low firing rates may actually be very common in the cortex (Henze et al., 2000; Shoham et al., 2006).

**Local Field Potentials**

The obvious drawback of single-unit recording is that it provides information mainly on the output of the recorded single neuron with no access to its subthreshold integrative processes or the associational operations taking place. To this end, we also record the LFP, to which these processes do contribute. LFPs are recorded when the impedance of the microelectrode is sufficiently low and its exposed tip is a bit farther from the spike generating sources, so that action potentials do not dominate the neural signal. The electrode then monitors the totality of the potentials. LFPs are related both to integrative processes (dendritic events) and to spikes generated by several hundreds of neurons (Lorente de No, 1947) and they represent mostly slow events reflecting cooperative activity in neural populations. They rather reflect the input of a given cortical area as well as its local intracortical processing, including the activity of excitatory and inhibitory interneurons. Based on current-source density (CSD) analysis and combined field potential and intracellular recordings, Mitzdorf (1985; 1987) suggested that LFPs reflect a weighted average of synchronized dendro-somatic components of the synaptic signals of a neural population near the electrode tip. Studies of inhibitory networks in the hippocampus (Buzsáki and Chrobak, 1995; Kandel and Buzsáki, 1997; Kocsis et al., 1999) have shown that other types of slow activity, including voltage-dependent membrane oscillations (Kamondi et al., 1998) and spike afterpotentials, also contribute to the LFP (Buzsáki et al., 1988). The soma-dendritic spikes in neurons of the CNS are generally followed by afterpotentials, a brief delayed afterdepolarization, and a longer lasting afterhyperpolarization (Granit et al., 1963; Gustafsson, 1984), which have a duration on the order of tens of ms (Kobayashi et al., 1997; Harada and Takahashi, 1983; Higashi et al., 1993).

Another finding in studies combining EEG and intracortical recording was that unlike MUA, the magnitude of the slow fluctuations was not correlated with cell size, but instead reflected the extent and geometry of dendrites at the recording site (Buchwald et al., 1965; Fromm and Bond, 1964, 1967) Cells in a so-called open field geometrical arrangement, in which dendrites extend in one direction and somata in another produce strong dendrite-to-soma dipoles when they are activated by synchronous synaptic input. The pyramidal cells with their apical dendrites running parallel to each other and perpendicular to the pial surface (Figure 1.2.6) form an ideal open field arrangement, and contribute maximally to both the macroscopically measured EEG and the LFP. But the dependence on geometry also implies that neurons that are oriented horizontally or have spherical symmetric dendritic fields (closed field arrangement) contribute less efficiently or not at all to the sum of potentials. Because of this large contributions arise from pyramidal/Purkinje neurons with interneurons often contributing less (cf. Chapter 1.1).

LFP and MUA signals can be separated by filtering; a high-pass filter with cutoff of approximately 500 Hz is typically used to obtain MUA and a low-pass filter cutoff of ~300 Hz to obtain the LFP. The modulations in the LFP are classified in a number of specific frequency bands initially introduced in the EEG literature. EEG is subdivided into frequency bands known as delta (DC–4 Hz), theta (4–8 Hz), alpha (8–12 Hz), beta (12–24 Hz) and gamma (24–40/80 Hz) (Lindsley and Wicke, 1974; Basar, 1980; Steriade, 1991; Steriade and Hobson, 1976). The classification is based on the strong correlation of each band with a distinct behavioral state. An alternative band separation is based on information theory, where the information carried by the different bands in the LFP or MUA range was calculated in recordings from monkeys that were viewing movies with natural images (Belitski et al., 2008). The most informative LFP frequency ranges were 1–8 Hz and 60–100 Hz. Positive signal correlations were found between LFPs (60–100 Hz) and spikes, and between the frequencies within the 60–100 Hz LFP range, suggesting that the 60–100 Hz LFP range and spikes are possibly generated within the same network. LFPs in the range of 20–60 Hz carried very little information about the stimulus, although they shared strong trial-to-trial correlations, indicating that they might be influenced by a common source such as diffuse neuromodulatory input.

**Spatial extent and propagation of neural signals**

The volume from which electrical signals are measured by a recording electrode depends on the properties of the
electrode. The activity from each point within the volume is weighted by a factor depending on the distance from the tip of the electrode. Single-unit activity or separable spikes are typically recorded from areas close to the electrode, for instance, Henze et al. (2000) and Gray et al. (1995) measured spikes within 50–150 μm from a tetrode. Electrodes with exposed tips of approximately 100 μm (impedance from 40–120 kΩ) were estimated to record from a sphere with a radius of 50–350 μm (Grover and Buchwald, 1970; Legatt et al., 1980; Nicholson and Linas, 1971). The volume from which LFP signals arise is larger, and the spatial extent of LFP summation can be calculated by computing the coherence of

---

**Figure 1.2.7.** Coherence between signal recorded at different electrodes as a function of electrode distance for the different neural signals. (A): diagram of the electrode array used in physiological experiments, a 4x4 electrode array with inter-electrode distance of 1 mm was used. (B): position of two electrodes (arrows) in combined MRI and physiology experiments. (C): Average of all fits to the coherence-over-distance data for the different neural signals. The plots show the drop-off of coherence with inter-electrode distance. The gray shading indicates the 1–99% confidence intervals. The distance at which the coherence is halved is marked by the red lines and shown next to the panels. The loss of coherence with distance was comparable for all neural frequency bands (1–3 mm). Adapted from Goense and Logothetis (2008), with permission.
LFP as a function of inter-electrode distance in experiments using simultaneous multi-electrode recordings (Juergens et al., 1996). The volume from which LFP signals arise is estimated to range from 0.25 to 3 mm distance from the electrode tip (Juergens et al., 1999; Katzner et al., 2009; Nauhaus et al., 2009; Mitzdorf, 1987).

The electrical properties of the conductive medium through which the current travels will obviously affect the voltages that are recorded. Within the physiological frequency range a current source can be described as a simple static point source. The description of the volume conductor is further simplified by the fact that the propagation of signals in gray matter is independent of their frequency. Intracranial measurements showed that the tissue impedance is actually frequency-independent and isotropic, allowing the description of the cortex as an ohmic resistor (Logothetis et al., 2007). Hence, this implies that the distance over which the signals can be measured is not dependent on the frequency of the signals but is determined by the relative size of the electrical sources.

The spatial extent of the different neural signals can be estimated by coherence analysis. The coherence between two electrodes is measured as a function of the distance between the electrodes for the different neural signals (Figure 1.2.7). This showed that the half-maximum of the coherence-to-distance functions was 2.9 mm for the 2–8 Hz LFP range, 2.4 mm for the 8–15 Hz range, 1.9 mm for the 20–60 Hz range, and 1.5 mm for MUA (Goense and Logothetis, 2008). Comparable distance-ranges were found for the LFP in V1/V2 (Juergens et al., 1999) and for the similarity of object preferences encoded by MUA and LFP in macaque IT (Kreiman et al., 2006). These findings suggest that the range over which the neural signals are measured in V1 are similar, and in the range of the typical voxel sizes in fMRI, and that this holds for the LFP band as well as for the MUA.

**Combined MRI and Physiology**

To address some of the questions raised earlier, we simultaneously measured fMRI and neurophysiological signals. Measurement of BOLD and neural signals is ideally done simultaneously; an important reason for this is that nonsimultaneous fMRI/electrophysiological recordings increases experimental variability, making subtle or not so subtle effects harder to discriminate. For instance, in anesthetized animals the response of the animal to anesthesia can be different from one animal to the next or from one experiment to the next. In awake subjects also, the magnitude of the BOLD response is different for different individuals and different sessions, and variability can also arise due to factors relating to the behavioral state of the animal, like attention, arousal, etc.

Given that the method of analysis is based on finding trends or correlations between phenomena that often yield weak signals, like the BOLD signal, or are highly correlated, like different neural signals, the added variance of nonsimultaneous measurement may further complicate interpretation of data. That said, to record BOLD and intracortical electrophysiological signals simultaneously some formidable hurdles need to be overcome. These are the interference of the scanner on the electrophysiological recordings and measuring equipment, and the possible degrading effect of the recording hardware on MR image quality. The scanner introduces electrical and mechanical interference on the electrophysiological signal due to the switching of the gradients (driven by 100–1000 V for our systems for example) and strong RF-signals for excitation (output of the amplifier in the kW-range) that interfere with the neurophysiological recording (10–100s of μV). The recording hardware can cause susceptibility artifacts in the images wherever metal is used, and any conducting leads into the magnet bore act as an antenna and can drag in noise. This is a problem because the received MR signal is a low voltage signal (μV-range) and hence image quality is sensitive to RF interference. In addition, all equipment that enters the bore needs to be nonmagnetic, and it should function in strong magnetic fields. The latter is a problem for equipment that enters the magnet bore, and for equipment that stays outside the bore it can be a problem if the magnet is not shielded, as is the case for instance for ultra-high-field magnets (Thulborn, 2006).

**Methodology and Practical Aspects**

We present a brief overview of our methodology for simultaneous fMRI and electrophysiological recordings. For details on the setup and experimental procedures we refer to the relevant publications (Goense and Logothetis, 2008; Keliris et al., 2007; Logothetis et al., 1999; Logothetis et al., 2001; Logothetis et al., 2002; Oeltermann et al., 2007; Pfeuffer et al., 2004). fMRI-eletrophysiology experiments are performed on vertical scanners (4.7T and 7T with 40 and 60 cm diameter bore respectively, BioSpec 47/40v and 70/60v, Bruker BioSpin, Ettlingen, Germany). These setups further consist of custom-built primate chairs and devices to fixate the monkey’s head, equipment to display the stimulus, and anesthesia monitoring equipment for anesthetized monkeys, and for awake monkeys: sensors to record the animal’s movement and track its eye-movements and a juice-delivery system.

Prior to experiments animals are implanted stereotaxically under general anesthesia with a custom-made headpost to fixate the animal’s head to the primate chair and a recording chamber (Logothetis et al., 2001; Logothetis et al., 2002). Animals used in awake experiments are extensively trained using operant conditioning methods on the behavioral task and to remain motionless for the duration of the experimental trial.

**Electrophysiological Recording**

For electrophysiological recording a small trepanation is made through which the electrode is introduced into the brain, and the electrode drive-assembly is fastened on the chamber with
the electrode positioned above the craniotomy. The drive-assembly consists of an electrode holder, a mechanism for manually advancing the electrode, and the magnetic-field sensor for the “near” interference circuit (see below) mounted on the microdrive (Figure 1.2.8). The electrode holder is composed of a glass-coated platinum-iridium electrode, the sensor for the “far” interference circuit (see below), and the ground contact. The holder consists of three concentric metallic cylinders: the inner cylinder is the contact point for the electrode, the middle the far-interference sensor, and the outer layer serves as the amplifier ground. These cylinders, in particular the outer cylinder, are a rotation-symmetric shield for the electrode, permitting optimal ground contact to the animal and avoiding loops susceptible to induction. Ground contact is provided by filling the chamber with deuterium saline and agar, which also reduces susceptibility artifacts in the MR images from air in the chamber.

The animal has a gold wire implanted (awake monkey) or a silver mouth electrode (anesthetized monkey) that acts as a “feedback” electrode. This electrode is part of the “far” interference compensation circuit and delivers the calculated counterinterference signal to the animal. The three-coil magnetic field sensor, necessary for the “near” interference compensation circuit, is positioned on or near the electrode-drive. Prior to the first functional scan the interference compensation is manually adjusted such that gradient interference on the electrode signal is minimal.

Signal Acquisition and Interference Compensation

The main problems arising when doing electrophysiology in the scanner are interference due to gradient switching and the fact that the preamplifier needs to be moved outside the magnet, which requires use of long cables that lead to large signal losses due to cable capacitance. To avoid signal loss, current is measured instead of voltage and prior to amplification the current is converted to voltage (see Oeltermann et al., 2007). The problem of interference was solved by using two separate interference compensation circuits (Logothetis et al., 2001; Oeltermann et al., 2007): a circuit to compensate for far interference arising from a distance larger than the distance from electrode tip to electrode ground and a circuit that compensates for near interference originating from the immediate vicinity of the electrode tip (Figure 1.2.9). Far interference arises due to capacitive coupling of the animal to metal resulting from the metal-to-electrolyte interface of the electrode. This allows interference currents to flow from, for example, ECG lines to the animal. By placing a sensor around the electrode and feeding an inverted copy of all interfering signals through it interference currents flowing to the animal can be eliminated (Logothetis et al., 2001; Oeltermann et al., 2007).

Because of the finite distance between the above sensor and electrode, interference originating from areas within this distance cannot be compensated for by the far-interference circuit. This interference is large enough to be problematic and hence it was compensated for by a near-interference circuit. Magnetic field changes due to the gradients were monitored by three small, orthogonally oriented coils positioned near the electrode, and the measured signal was added to the “ground” of the current-to-voltage converter to neutralize the interference.
small coils are oriented orthogonally and the gain and sign of the signals is manually adjustable, it is possible to simulate the induction voltage in a wire loop of any diameter and orientation. In this way a virtual wire loop that has opposite direction of winding can be adjusted such that the loops caused by asymmetries in the electrode holder and cable are effectively compensated (Logothetis et al., 2001; Oeltermann et al., 2007).

The above interference reduction techniques yield a non-saturated, measurable signal that, however, still contains some gradient interference. This residual interference is eliminated by principal component analysis (PCA) and elimination of those principal components that strongly correlate with the interference directly recorded from the gradient amplifiers results in a “clean” signal (Logothetis et al., 2001).

Neural Basis of the BOLD Response

To investigate the neural origins of the fMRI response and the coupling between neural and BOLD signals, simultaneous fMRI and electrophysiology were performed in anesthetized (Logothetis et al., 2001; Rauch et al., 2008; Shmuel et al., 2006) and awake monkeys (Goense and Logothetis, 2008). In the well-known 2001 study it was shown that the LFP is generally a better predictor of the BOLD response than the MUA (Logothetis et al., 2001). Figure 1.2.10 shows an example of a comparison of the time course of the BOLD signal and the neural signals in the MUA- and LFP bands in an awake monkey.

The figure shows time courses of seven band-limited power (BLP) signals extracted from the comprehensive neurophysiological signal following removal of gradient interference, band separation, and rectification. The first

\[
C_e, C_s, C_g: \text{electrode, sensor, ground capacitance}
\]
three bands are known from the EEG literature, while the other bands were defined based on recent work (Belitski et al., 2008). There the relationship between visual information carried by different frequency bands of LFP and spikes was investigated in recordings while the monkey was viewing five-minute color movie clips. This ensured that the stimulation was diverse and likely stimulated all visual cortices not affected by anesthesia.

Figure 1.2.10. Dependencies between BOLD and neural signals in V1 in awake monkeys. (A) Band-separated neural signals (black) in response to a $6^\circ$ visual stimulus (top right). The band-limited power (BLP) signals were convolved with a theoretical hemodynamic response (HRF) and used as regressors for each band (red). Green shading indicates the times the stimulus was presented. The response to the stimulus is especially pronounced in the neuromodulatory and gamma bands. (B) Functional activation map superimposed on anatomical images. The location of the electrode is shown by the arrow. (C) The BOLD time course acquired at a temporal resolution of 250 ms shows obvious modulation to the stimulus. The output of the GLM analysis and F-test yielded significant $p$-values for all neural bands (D) indicating that all bands contributed significantly to the BOLD response. (E) Beta values lacked dramatic differences across bands. Adapted from Goense and Logothetis (2008), with permission.
In Figure 1.2.10, however, the stimuli were simple geometrical shapes that optimally drive V1. Moreover, due to the limitations imposed by the behavioral task, the stimuli could only be shown for short periods of time in sequential trials. It is therefore not surprising that the time courses of the signals in all bands are relatively similar to each other due to the presence of onset responses. Nonetheless, analysis using the general linear model (GLM) (Friston et al., 1995) revealed a differential contribution of the different neural signals to the BOLD response. Whether a frequency band has a unique contribution to the BOLD signal can be assessed by calculating the F-ratio, which showed that all frequency bands contributed significantly to the BOLD signal, and significant F-ratios for all bands indicates that each band explains a component of the BOLD response that cannot be explained by any of the other bands (Goense and Logothetis, 2008). The average beta values were comparable across frequency bands, suggesting that under these stimulus conditions no one single band especially determines the BOLD response.

Studies in awake and anesthetized monkeys have shown higher correlation coefficients between LFP and the BOLD signal than between MUA and the BOLD signal (Goense and Logothetis, 2008; Logothetis et al., 2001). This implies that the overall synaptic activity or the input of an area is a stronger generator of the BOLD signal than its output. But although the correlation of the LFP to the BOLD signal is consistently higher than the correlation of the MUA to the BOLD signal, MUA is also positively correlated, and is also significant. High correlations of spiking with the BOLD response were also found in humans (Mukamel et al., 2005). This is not surprising given that in most cases the MUA is correlated to the LFP, and in most cases a positive correlation exists between the input and the output of a neural system. Thus, based on differences in correlation coefficients we cannot unambiguously determine which signal better drives the BOLD response. This is further complicated by the fact that the neural and BOLD responses are often noisy. Because of the correlation between MUA and the BOLD signal, one needs to find circuits or stimulus conditions where

**Figure 1.2.11.** Dissociation between the MUA and BOLD response. (A) Functional activation maps in response to a 6° peripheral rotating checkerboard stimulus centered at 6.5° (inset in B). The arrow indicates the location of the electrode. (B) The average MUA, LFP in the 20–60 Hz range and BOLD time courses show that while the neuromodulatory component of the LFP stayed elevated for the duration of the stimulus, the MUA rapidly returned to baseline after a transient onset response. The prolonged time course of the BOLD response suggests a more sustained driving mechanism of the BOLD response as opposed to the transient MUA signal. The dotted line shows the regressor, i.e., the neural signal convolved with the theoretical HRF, which indicates that the MUA-derived regressor cannot capture the sustained part of the BOLD response. Adapted from Goense and Logothetis (2008), with permission.
there is dissociation to obtain more conclusive evidence. In other words, conditions where the LFP is not or only weakly correlated with the MUA.

An example of such a case in the awake monkey is shown in Figure 1.2.11. Dissociation of the LFP and MUA can occur when there is strong adaptation and the LFP stays elevated long after the MUA has returned to baseline (Logothetis et al., 2001). In these cases the BOLD response also stayed elevated and this clearly shows the better correlation between BOLD and LFP than between BOLD and MUA. Similar results, i.e., better coupling of the hemodynamic response to the LFP in cases of dissociation, were observed by other groups, for instance in V1 of cats using optical imaging (Niessing et al., 2005). Another way to induce dissociation was demonstrated by Viswanathan and Freeman (2007), who used high temporal frequency stimuli that did not elicit spikes in V1 and who observed that LFP activity elicits changes in tissue oxygen in the absence of spiking.

Dissociation of LFP and MUA can also be induced by pharmacological intervention. For example intracortical injection of serotonin or a 5-HT1A serotonin receptor agonist abolishes spiking of the output neurons that are typically recorded in the MUA (Logothetis, 2003; Rauch et al., 2008). The LFP on the other hand did not change substantially, and similarly the BOLD response did not show changes. Such cases indicate that the signals recorded in the LFP and representing the input in an area are more likely driving the BOLD signal than the MUA signal, which mostly represents the output of large pyramidal neurons.

Other cases of dissociation of neural activity and metabolism or blood flow have been observed in structures where the anatomical and functional properties of the neural circuit allow a clear segregation between input and output. For instance, in the lateral superior olive (LSO) of the cat auditory system Nudo and Masterton (1986) observed that inhibitory synaptic activity increased 2-DG labeling although postsynaptic spiking is suppressed. In the cerebellum, Mathiesen et al. (1998) observed that stimulation of the parallel fiber system provides inhibitory synaptic input and inhibits Purkinje cell firing but that the CBF response remained.

Given the similar volumes from which the neural signals are recorded (see above) the stronger coupling of LFP to BOLD (Goense and Logothetis, 2008; Logothetis et al., 2001) can not be explained by differences in spatial summation of the neural signals. Also the spatial area from which the neural signals are sampled provides justification to use fMRI resolutions of 1–2 mm to determine the correlation between MR- and neural signals.

**LFP, Spikes, Metabolism, and Blood Flow**

The results above show that the BOLD response is better correlated with LFP than with MUA, implying that the BOLD response is better correlated with the input and local processing of the neurons in an area than with the output of the large pyramidal cells. Note that “synaptic input” does not necessarily imply input from another cortical or subcortical area, but it also refers to local (intrinsic) connections. Like the positive BOLD response, negative BOLD responses were also associated with decreases in LFP and MUA (Shmuel et al., 2006).

The energy demands of neurotransmission and spiking determine the blood flow to an activated area. Ultimately, the function of perfusion is the supply of nutrients and O₂, removal of waste products, and removal of heat. This raises questions about which processes have the highest metabolic demands (synaptic processes vs. spiking) and whether the neurovascular response is driven by the processes that have the highest metabolic demands or if there is some kind of anticipatory or feedforward process.

The brain consumes 20% of the body’s total energy (Sokoloff, 1969), and it is oxidative phosphorylation that feeds the brain. Most of the energy (50–80%) is used for neural signaling, while cellular maintenance processes like protein synthesis use a minor (5–15%) fraction of the energy budget (Ames III, 2000; Raichle and Mintun, 2006; Shulman et al., 2004; Riera et al., 2008). Because the sodium-potassium pump or Na⁺/K⁺-ATPase maintains the transmembrane electrochemical gradients that are the driving force for most signaling processes, the main energy consumer in the brain is the Na⁺/K⁺-ATPase (Sokoloff, 1999). The pump depends on oxidative phosphorylation for its energy needs (Erecinska and Silver, 1994) and high concentrations of Na⁺/K⁺-ATPase co-localize with high levels of cytochrome oxidase (CytOx) reactivity (Hevner et al., 1992; Wong-Riley, 1989). Cytochrome oxidase is the enzyme in the electron transport chain that catalyzes the reduction of O₂ to H₂O and hence is a marker for oxidative metabolism. The vascularization of the cortex also reflects its energy use: areas that have high energy needs have denser vascularization. For instance, in V1 energy use is highest in layer IV and it is higher in blobs than in interblobs. This is evidenced by higher CytOx-reactivity (Wong-Riley, 1989) and higher ¹⁴C-deoxyglucose (2-[¹⁴C]-DG) uptake (Kennedy et al., 1976), and is reflected in the higher vascularization of these areas (Weber et al., 2008).

At the cellular level, we can also ask which signaling processes use the most energy, for instance, spiking, synaptic transmission, neurotransmitter recycling, or the maintenance of transmembrane gradients (Ames III, 2000; Attwell and Laughlin, 2001). It is generally believed that synaptic transmission and associated processes use more energy (Ames III, 2000; Shulman et al., 2004), although the relative contribution of spiking is still debated (Attwell and Laughlin, 2001; Lennie, 2003). The higher presence of cytochrome oxidase and mitochondria in dendrites (especially in postsynaptic areas) compared to cell bodies and axons indicates that these are the more metabolically active sites (Wong-Riley, 1989). Similarly, Schwartz et al. (1979) also found that nerve terminals have a higher uptake of radiolabeled glucose than areas with cell bodies.
Despite these energy needs however, the cerebral metabolic rate of oxygen consumption, or CMRO$_2$, measured with PET showed little increase under visual stimulation, while blood flow and glucose consumption did increase, indicating an uncoupling between blood flow and oxidative metabolism (Fox and Raichle, 1986; Fox et al., 1988; Raichle and Mintun, 2006). These observations have incited the debate about the importance of glycolytic versus oxidative metabolism in brain function (Pellerin and Magistretti, 2003; Pellerin et al., 2007; Chih and Roberts Jr., 2003; Shulman et al., 2004; Gladden, 2004) as it was also found with MR spectroscopy that excess lactate is produced during functional activation (Mangia et al., 2007; Prichard et al., 1991; Sappey-Marinier et al., 1992).

Pellerin and Magistretti (1994; Pellerin et al., 2007) put forward the hypothesis that glutamate released at the synapse is taken up into astrocytes, where it induces glycolysis leading to the production of lactate, which is then released and taken up by neurons to be used as an energy substrate. This is also called the astrocyte-neuron lactate shuttle hypothesis (ANLSH). The competing view holds that both neurons and astrocytes use glucose as their main substrate, and lactate is produced when glycolysis exceeds the rate of oxidative metabolism, a situation that is both transient and potentially detrimental (Chih and Roberts Jr., 2003). Hence the debate focuses on the compartmentalization of glycolysis versus oxidative metabolism, with one side arguing that astrocyte metabolism is predominantly glycolytic and neural metabolism mostly oxidative and the other side that both oxidative and glycolytic metabolism occur in neurons and astrocytes.

Astrocytes have many different functions in the brain and play an important role in neurovascular coupling and metabolism. They are a subtype of glial cells characterized by their star shape and until quite recently they were mostly considered filler material. They are involved in regulating homeostasis in the brain (Simard and Nedergaard, 2004) and providing energy to the neurons by supplying nutrients to the neurons. Another function of astrocytes is the uptake of neurotransmitters released from nerve terminals, however, the neurons. Another function of astrocytes is the uptake of neurotransmitters released from nerve terminals, but it was observed that CMRO$_2$ increased with prolonged stimulation (Gjedde and Marrett, 2001; Mintun et al., 2002; Vlassenko et al., 2006b). Hence there may be a shift from initially more glycolytic metabolism toward more oxidative metabolism under sustained stimulation (Gjedde and Marrett, 2001; Mintun et al., 2002; Raichle and Mintun, 2006; Vlassenko et al., 2006a). However, further studies are needed to clarify the contributions of glycolysis and oxidative metabolism during both baseline and activation.

But whether metabolism under activation is primarily oxidative or glycolytic, or whether the nutrient that needs to be supplied is glucose or O$_2$ may not be the main factor in understanding the BOLD response. What matters for the BOLD signal is what drives the signaling cascade that leads to the increase in blood flow, given that it is the blood flow response that is taken as an indicator of increased neural activity. Hence, to understand the BOLD signal, we need to look at neurovascular coupling and ask what is the trigger for the neurovascular response. Again, the processes related to synaptic function—synaptic transmission, restoration of electrical gradients, and neurotransmitter recycling—elicit the functional blood flow increases (Wang et al., 2005; Lauritzen, 2005; Hoffmeyer et al., 2007; Iadecola and Nedergaard, 2007). Especially glutamatergic neurotransmission leads to an increase in blood flow (Li and Iadecola, 1994; Hoffmeyer et al., 2007; Gsell et al., 2006; Yang et al., 2003; Mathiesen et al., 1998) although there are numerous vasoactive signaling molecules and multiple vasodilatory and constrictive mechanisms (Cauli et al., 2004; Iadecola and Nedergaard, 2007; Iadecola, 2004). Molecules involved in signaling pathways leading to vasodilation during functional activation are also associated more with synaptic signaling (Zhang and Wong-Riley, 1996; Hoffmeyer et al., 2007; Yang et al., 2003) than with action potentials or metabolic signals (Attwell and Iadecola, 2002; Iadecola, 2004) although the redox state during activation possibly plays a role through the cytosolic NADH/NAD$^+$ in astrocytes (Ido et al., 2001; Ido et al., 2004; Raichle and Mintun, 2006; Vlassenko et al., 2006b). Many of these neurovascular coupling processes are also mediated by astrocytes, because of their key location between vasculature and neurons (Nedergaard et al., 2003). Their numerous processes contact the blood vessels and envelope synapses, and they are involved in the regulation of local blood flow, and hence are an important mediator of the BOLD signal. Their effect on the blood flow responses are again often triggered by neuronal glutamate release (Zonta et al., 2003; Takano et al., 2006; Schummers et al., 2008; Metea and Newman, 2006; Koehler et al., 2006; Petzold et al., 2008).

**The Cortical Circuit and the BOLD Response**

The BOLD response is associated more with input and local processing of cortical neurons than with their spiking output. But this opens further questions about what specific cortical processes and what properties of the cortical circuitry determine the BOLD response. For a better interpretation of fMRI it is essential that we understand which processes are represented in the BOLD response, and which less, or not at all
In essence this boils down to a question about the transfer function: Do all neural events contribute equally to the BOLD response, or do some events contribute more than others? The same questions can be asked for the LFP and metabolic signals.

Different cortical processes can and without much doubt do have different contributions to the BOLD, LFP, and MUA signals. The fMRI signal may also reflect different aspects of neural processing that are not always observed with single-unit recording. For instance, a binocular rivalry stimulus showed robust functional activation in V1, while only a small fraction of the single neuron’s responses were modulated by the percepts (Blake and Logothetis, 2002). Different processes, like feedforward vs. feedback processes, or stimulus driven or neuromodulatory processes, or for instance subcortical input vs. cortico-cortical input or recurrent intracortical input, can affect these signals differently. If we look at connectivity we find that most cortical connections are highly local. In contrast, subcortical input is rather weak in terms of number of synapses (Douglas and Martin, 2007; Peters and Payne, 1993; Peters et al., 1994), for instance thalamocortical input typically comprises only about 10–20% of synapses (Douglas and Martin, 2007) and in V1 only ~5% (Peters and Payne, 1993; Peters et al., 1994). It is not known whether the BOLD signal is more or less strongly weighted toward intracortical processing or cortico-cortical or thalamocortical processing.

Furthermore, different types of neurons may have different contributions to the BOLD signals than they do to the neural signals. For instance, interneurons are less visible in LFP and MUA signals than pyramidal neurons, but because they can have high firing rates they could possibly have substantial metabolic demands (Buzsaki et al., 2007) and a considerable effect on the BOLD response. Interneurons have also been shown to cause dilation and constriction of microvessels (Cauli et al., 2004). Another example is smaller cells that can also have higher firing rates and are more easily stimulated than large cells (Gur et al., 1999) and hence may have a substantial contribution to the BOLD response, although their contribution to the recorded MUA and LFP will be less.

Another question is how excitation and inhibition contribute to the fMRI responses (Logothetis, 2008; Buzsaki et al., 2007) and whether the BOLD signal is mediated by specific neurotransmitters or receptors. We can ask the same question about the LFP and MUA because different types of neurotransmission may also not be equally represented in the recorded neural signals. Glutamatergic excitatory neurotransmission is commonly associated with BOLD and CBF responses (Gsell et al., 2006; Hoffmeyer et al., 2007). One reason is that these are the most common synapses in the cortex and they outnumber inhibitory synapses by about five to one (Braitenberg and Schüz, 1998). Glutamatergic excitatory synapses are also correlated with high cytochrome oxidase levels (Wong-Riley, 1989). However, although N-methyl-D-aspartate (NMDA) and α-amino-3-hydroxy-5-methyl-4-isoxazole propionate (AMPA) receptor activity were both shown to contribute to the CBF and BOLD responses, in contrast to AMPA receptor activity, NMDA receptor activity did not contribute to the LFP (Mathiesen et al., 1998; Gsell et al., 2006; Hoffmeyer et al., 2007). NMDA receptor activity is linked to increased blood flow through the nitric oxide (NO) signaling pathway (Li and Iadecola, 1994; Akgoren et al., 1994) which mediates glutamate induced blood flow increases. For instance nitric oxide synthase (NOS) knockout mice showed decreased activation-induced blood flow responses in the cerebellum (Yang et al., 2003). The effect of inhibitory neurotransmission on the blood flow responses is much less clear (Logothetis, 2008; Buzsaki et al., 2007) and may be small (Waldvogel et al., 2000), which could be due to the lower number of inhibitory synapses. However a contribution of inhibitory neurotransmission to the BOLD signal in the cortex is not unlikely because gamma aminobutyric acid (GABA)ergic neurotransmission has been shown to account for ~15% of the energy consumption (Patel et al., 2005; Hyder et al., 2006). Furthermore, inhibitory neurons can have high firing rates, and their synapses are also associated with mitochondria and high CytOx levels (Wong-Riley, 1989). Inhibition has also been shown to increase 2-DG uptake (McCasland and Hibbard, 1997; Nudo and Masterton, 1986; Ackermann et al., 1984), and GABA was shown to induce vasodilation in hippocampal microvessels (Fergus and Lee, 1997). Finally, other neurotransmitters and neuromodulators are known to be vasoactive and may play a role in the BOLD response, for example acetylcholine, serotonin, dopamine, noradrenaline (Attwell and Iadecola, 2002).

Hence, although it is without doubt true that the BOLD signal is related to neural processing, what cortical processes exactly it does and does not represent is still far from clear. For many studies, knowing that activation is due to neural activity may be all that one needs to know. But if we want to go beyond that and interpret the BOLD findings in the context of these different neural processes it becomes increasingly important to better understand the neurovascular response and its relation to the underlying neural processes.

**Conclusion**

The finding that the fMRI response is better correlated with the LFP, which represents the input and local processing of a neural circuit, than to the MUA, representing the output, has important implications for the interpretation of the BOLD signal. Especially when BOLD signals are compared to, or interpreted in terms of, the results obtained in electrophysiological (typically single-unit) recordings. Although in most cases spiking output is correlated with synaptic input, and hence in most cases the spiking activity and BOLD are well correlated, there may be cortical processes where the

...
relationship and thus the interpretation of the BOLD signal is not so straightforward.

Although we now have a clearer picture of the origins of the fMRI responses, we still do not fully understand the details of the processes underlying the BOLD response, and many issues are still unresolved. Hence the obvious conclusion is that further experiments are needed to elucidate the exact nature of the functional responses. This of course needs a better understanding of the metabolic and neurovascular processes that take place during functional activation, regardless of whether these are measured by MRI, PET, or optical methods. And although the neurophysiological methods like single- and multi-unit recording and EEG have been around for decades, many questions of what exactly is being measured are still unresolved. Progress can be made by improving current functional imaging methods or development of new functional imaging methodology, the combination of multisite or array electrode recordings with functional imaging, or intervention by for instance pharmacological injections (Kid et al., 2006; Stefanovic et al., 2007; Rauch et al., 2008) or electrical stimulation (Tolias et al., 2005). Many insights in recent years have come from a combination of methods, not only fMRI or spectroscopy and neurophysiology (Logothetis et al., 2001; Mukamel et al., 2005; Smith et al., 2002), but also from combining optical methods with electrophysiology (Lauritzen and Gold, 2003; Arieli and Grinvald, 2002; Arieli et al., 1996) or fMRI (Kennerley et al., 2005). Because the different methods have different strengths and weaknesses, they provide complementary information, underscoring the importance of multidisciplinary work.

**ACKNOWLEDGMENTS**

We wish to thank Dr. Alexander Rauch and Dr. Kevin Whittingstall for comments on an earlier version of the manuscript. Supported by the Max-Planck Society.

**REFERENCES**


