TWO DECADES OF FUNCTIONAL IMAGING: FROM NUCLEAR SPINS TO CORTICAL COLUMNS

Kâmil Uğurbil
Center for Magnetic Resonance Research
University of Minnesota, Minneapolis, Minnesota

EMAIL: kamil@cmrr.umn.edu

Accepted December 31, 2011

KEYWORDS
fmri, high field, ultrahigh field, magnetic resonance

ABSTRACT
Since its introduction in 1992, there has been a revolution in the ability to image brain function with functional magnetic resonance (fMRI), going from early experiments demonstrating relatively course images of activity in the visual cortex to mapping cortical columns and to “brain reading” that constructs mental experiences of an individual, all using the fact that we were endowed with a complex paramagnetic molecule sequestered in our blood vessels and that neuronal activity has spatially-specific metabolic and physiologic consequences. These two decades of fMRI is marked by incessant improvements in instrumentation, innovative developments in image acquisition and reconstruction methods, and a significant expansion in our knowledge of neurovascular coupling. Collectively, this body of work has brought us recently to the point of depicting functional activity in three dimensions, in the entire human brain with submillimeter resolution. Some aspects of these accomplishments and the rational for their pursuit is reviewed and presented together with a personal history of the development of fMRI.
INTRODUCTION

In 2012, functional magnetic resonance imaging (fMRI) based on blood oxygen level dependent (BOLD) contrast will be two decades old. Because of this two-decade milestone, there has been a flurry of activity to publish historical accounts and personal perspectives. In this spirit, I have decided to include a personal perspective in this article intermixed with a review of fMRI because functional imaging has been a central research focus since its discovery in the laboratory that I lead, the Center for Magnetic Resonance (CMRR) at the University of Minnesota. For this laboratory, the development of fMRI also coincides with and is intertwined with the development of high (4 Tesla) and ultrahigh fields (7 Tesla and above) for magnetic resonance (MR) imaging. As such this personal perspective also includes a history of high and ultrahigh field MR.

Since its introduction in 1992 (Bandettini et al. 1992; Kwong et al. 1992; Ogawa et al. 1992), there has been a revolution in the ability to image brain function, going from early experiments demonstrating relatively course images of activity in the visual cortex, to mapping cortical columns, and to brain reading that constructs mental experiences of an individual, all using the fact that we were
endowed with a complex paramagnetic molecule sequestered in our blood vessels and that neuronal activity has spatially-specific metabolic and physiologic consequences. Today, it is possible to depict functional activity in three dimensions in the whole brain with approximately a millimeter resolution, and image with even higher spatial resolution at the level of cortical columns in restricted volumes (Menon et al. 1997; Cheng et al. 2001; Goodyear and Menon 2001; Fukuda et al. 2006; Sun et al. 2007; Yacoub et al. 2007; Yacoub et al. 2008; Chaimow et al. 2011; Shmuel et al. 2010; Zimmermann et al. 2011). An example of the latter includes orientation columns that were never previously visualized in the human brain but were successfully mapped with fMRI at 7T (Yacoub et al. 2008) in the primary visual area V1 together with ocular dominance columns (ODC), revealing the organizational relationship between them (Yacoub et al. 2008) (Figure 1). While a lot has been known about human ODCs for quite some time due to the ability to image them even in the postmortem brain, orientation columns require functional methods for their depiction and were never imaged in humans before. It was possible to do so only with a major effort to push fMRI to ultrahigh magnetic fields, develop new image acquisition strategies, and subsequent to extensive theoretical and experimental studies that made significant inroads into understanding the complex set of mechanisms underlying the coupling between neuronal activity and MRI signals. This combined effort was indispensable for attaining improvements in spatial accuracy, functional contrast to noise ratio (fCNR), and intrinsic image signal-to-noise ratio. This review focuses on these methodological developments that ultimately enabled some of the most advanced applications of fMRI in the human brain.

DEVELOPMENT OF FMRI

The first functional images of the human brain were obtained using an exogenous, gadolinium based contrast agent that was at the time (and still is) available for clinical use (Belliveau et al. 1991). Although this was an exciting development, the approach has many shortcomings: It required two bolus injections of contrast agent, separated by many minutes, for a single functional map, limited amount of time for data acquisition during the transient bolus passage through the brain, and a limited number of possible repeated
FIGURE 1: Functional maps of ocular dominance columns (a, b) and orientation columns (c, d) in the human brain obtained with SE fMRI at 7 Tesla in two subject. The black lines in (c), and (d) depict the boundaries of the ocular dominance columns seen in (a), and (b), respectively. The black and white dots in c and d identify the centers of clockwise and counterclockwise rotating pinwheels with each color representing an orientation. The white bars (lower left hand corner in (c) and (d) designate 1 mm scale. From (Yacoub et al. 2008).

Contrast agent injections due to accumulation of the contrast agent in the blood and, hence, the saturation of relaxation enhancement with additional injections. Ultimately, human applications of this technique would have been limited by safety concerns because gadolinium-containing contrast agents have become a suspected causal factor for nephrogenic systemic fibrosis (NSF) (e.g. Marckmann et al. 2006; Kuo et al. 2007). Because of all these reasons, it would not have been possible with this technique to perform all the incredibly information-rich and complex fMRI studies that are feasible today, and the impact on understanding brain function would have been quite modest. Fortuitously, the exogenous contrast agent based
functional imaging approach was rapidly followed by what is now often referred to as BOLD fMRI. This technique relies on intrinsic contrast mechanisms associated with deoxyhemoglobin changes and, as such, does not suffer from the shortcomings of the exogenous contrast based method. Clearly, it is this latter development that revolutionized the study of brain function. We at CMRR at the University of Minnesota were one of two groups that independently and simultaneously initiated and conducted the experiments that introduced fMRI (Ogawa et al. 1992), the other being the group at Massachusetts General Hospital (MGH) (Kwong et al. 1992), an effort that is largely credited to Ken Kwong. A third group from the Medical College of Wisconsin, Milwaukee also reported results of functional imaging in the human brain the same year (Bandettini et al. 1992). However, this group acknowledged being motivated by results from the MGH effort.

The fMRI in CMRR came about because of the work Seiji Ogawa did at Bell Labs, introducing the BOLD effect (Ogawa et al. 1990; Ogawa et al. 1990; Ogawa and Lee 1990). These early experiments conducted on rats were not directly on functional imaging and did not show functional mapping; rather, they demonstrated that metabolic perturbations such as hypoglycemia and graded levels of oxygen in the inhaled gas mixture affected the visibility of venous blood vessels. These perturbations altered the visibility of intracortical blood vessels that were detectable in gradient recalled echo (GRE) images of the rat brain where a delay TE is inserted between the excitation of spins and their subsequent detection to generate an image. This effect was ascribed to intravoxel magnetic field inhomogeneity induced by magnetic susceptibility differences between the blood vessel interior and the extravascular space; the origin of the susceptibility gradient itself was assigned to the presence of paramagnetic deoxyhemoglobin compartmentalized within the blood vessels. Therefore, in MR images sensitized to magnetic field inhomogeneities, such as simple GRE images with a TE delay, signal intensities are altered if the regional deoxyhemoglobin (dHb) content is perturbed. Thus, the BOLD effect, as described was an extravascular effect impacting water molecules in the surrounding tissue and cerebrospinal fluid (CSF).

Although the original BOLD effect was demonstrated with metabolic and physiologic perturbations, the dHb changes that are the ultimate source of the BOLD effect can occur with alterations in brain activity because of spatially specific metabolic and hemo-
dynamic responses to enhanced neuronal activity. It was earlier suggested that in response to stimulus induced neuronal activity, regional cerebral blood flow (CBF) increases while the cerebral metabolic rate of oxygen consumption (CMRO$_2$) in the same area is not elevated commensurably (Fox and Raichle 1986; Raichle 1987; Fox et al. 1988), implying that with increased neuronal activity, O$_2$ extraction fraction, and subsequently dHb content per unit volume of brain tissue would decrease. The regionally specific CBF changes coupled to brain activity were well known and accepted at the time fMRI experiments were being attempted; this phenomenon was in fact used for the very early human functional imaging experiments performed with xenon, and later by positron emission tomography (PET) (Raichle 2009). It was also well known at the time that these activity-induced CBF increases are also accompanied by commensurate elevations in glucose consumption rate. However, the concept that CMRO$_2$ did not increase equivalently as CBF was controversial at the time, in part because it was quite counterintuitive given the extreme aerobic capacity of the brain and, in part, because the PET measurements of CMRO$_2$ are complex.

Nevertheless, if the cerebral dHb content changes with altered neuronal activity induced by a stimulus or a task relative to a basal state, so will the BOLD effect in regions of the brain engaged in this task or stimulus, leading to a functional brain image (i.e. to fMRI). This served as the motivation to explore functional imaging using the very simple strategy of collecting gradient recalled echo images with a delay between signal excitation and reception, expecting the BOLD effect to provide functional contrast.

I became aware of Ogawa’s original work on the BOLD contrast from various conferences which we both attended. I had seen some of his presentations (e.g. the 1999 annual meeting of the Society of Magnetic Resonance in Medicine (SMRM) in Amsterdam), and also discussed this work informally when we met. Seiji Ogawa and I knew each other well; we were colleagues that had worked together for several years in the same group at the Bell Laboratories, focused on establishing in vivo applications of the magnetic resonance phenomenon. The link to oxygen metabolism, at the time my research focus, albeit in the heart and not in the brain, is what initially attracted my attention to the BOLD effect. Although the original description of the BOLD contrast was based on manipulation of blood oxygenation levels in the rat model through pharmaceutical and/or metabolic interventions, in his paper published in the Proceedings
of the National Academy of Sciences, USA (PNAS), submitted in August 1990 and published in December 1990 (Ogawa et al. 1990), Ogawa and his coworkers suggested the use of this method possibly to achieve functional imaging in the brain, in a way analogous to the PET approach but, unlike PET, using an endogenous contrast mechanism instead.

The original work on the BOLD effect was carried out in the late 1980s. At about the same time, we at CMRR were attempting to establish a human imaging and spectroscopy capability based on 4 Tesla magnetic field strength, at a time when the magnetic field used in high-end instruments was 1.5 Tesla. The 4 Tesla system provided an opportunity to pursue unique experiments that had the potential to provide distinctive results, especially using imaging mechanisms based on magnetic susceptibility. In this context, Ogawa and I started discussing the possibility of working together to achieve functional imaging in the human brain using the BOLD contrast and the 4 Tesla system (this was even before the Ogawa et al. 1990 PNAS paper was submitted). We were simply waiting for this instrument to arrive and become adequately operational before we started the fMRI experiments. Had the 4T system been delivered earlier or had it functioned right away we would have achieved fMRI earlier.

As we waited for the 4T to become operational, we did not think of pursuing fMRI at 1.5T because we were focused on the susceptibility mechanism of the BOLD contrast. Magnetic field perturbations in a sample, caused by the presence of boundaries between compartments of different magnetic susceptibility, increase linearly with the external magnetic field applied. The original rat experiments on the BOLD effect were performed mainly at 7 Tesla, and to a lesser extent at 4.7 Tesla. The latter was close to the magnetic field we were going to establish for human studies. As such, we thought 4 Tesla would be adequate to detect brain function with the BOLD mechanism, if deoxyhemoglobin changes coupled to neuronal activity were as large as predicted from the existing PET data of the time. We did not think BOLD fMRI would work at low fields like 1.5T. In principle, we were right. Today, we know that the fMRI signal is quite complex and has numerous contributions (discussed in greater detail later on); such contributions include inflow effects that are mostly associated with flow increases in large blood vessels (e.g. Duyn et al. 1994; Segebarth et al. 1994) and that are field independent, intravascular effects due to changes in blood $T_2$ and $T_2^*$ (e.g. Duong et al. 2003; Silvennoinen et al. 2003) that actually
are prominent at lower magnetic fields (e.g. 1.5 Tesla), and signal changes associated with large veins with diameters that are of the order of the voxel size (e.g. Hoogenraad et al. 1999), again an effect that is prominent at lower fields such as 1.5T. Extravascular BOLD effect arising from intracortical veins and capillaries is relatively small, if present at all, at 1.5T (e.g. Hoogenraad et al. 2001), but becomes significant at high and ultrahigh fields. As such, 1.5T fMRI data are mostly associated with large veins but because they are acquired at low resolution, they appear spread out, looking similar to what is seen in PET functional images. Nevertheless, for many questions these functional maps provide adequate resolution and have produced useful results.

When the 4 Tesla instrument finally did arrive in 1990, much had to be done to bring it to an operational state since this system was not a well-developed platform. In addition, the magnet was damaged during transport, and had to be repaired, causing a significant delay. Echo Planar Imaging (EPI) that has now become the most commonly employed imaging approach for fMRI was not available generally on any system, let alone a 4 Tesla system. We used gradient recalled echo imaging (i.e. FLASH). We were collecting data sometime early in 1991. Obviously, in these early experiments, we worried if the results were real, if they were motion artifacts, or instrumental glitches etc. We also expected the stimulation-evoked signal to be a decrease relative to the baseline state, being prejudiced with familiarity of strongly coupled oxygen consumption increases associated with work and metabolism in other systems. We were aware of the Fox and Raichle paper reporting minimal to no oxygen consumption change with visual stimulation studies with PET-based measurements of oxygen consumption (Fox et al. 1988), but, as many others in the field, were skeptical of it. When we started to be convinced that we were seeing stimulus-induced positive signal changes due to visual stimulation, we had to acknowledge that at least qualitatively, the PET data were correct after all.

By the time we went to the Society of Magnetic Resonance in Medicine (SMRM) annual meeting that was held in San Francisco in August 1991, we had functional images and so did MGH. I think, however, that the MGH colleagues and us were likely being very cautious since this was such new territory and signal changes observed so small. Otherwise both groups would have rushed to publish these unique results. Yet it took more or less another six months
or so before the papers from these two groups were submitted for publication within five days of each other.

We submitted our paper to Nature. It was rejected without being sent for scientific review, with the usual rejection letter saying that it was not of *general interest*. After the rejection, we recouped and sent it to Proceedings of the National Academy of Sciences, USA (PNAS) where it was received in March 1992 and appeared in press in July 1992 (Ogawa et al. 1992).

Since the introduction of BOLD based fMRI, other MRI techniques have also been developed for functional brain mapping. It is now also possible to generate functional maps based on relative or absolute measures of CBF change (e.g. Kwong et al. 1992; Edelman et al. 1994; Kim 1995; Kim and Tsekos 1997; Wong et al. 1998; Wong et al. 2006; Wu and Wong 2006; Wu and Wong 2007; Pohmann et al. 2010) or change in the cerebral blood volume (CBV) (Lu et al. 2003; Lu et al. 2005) coupled to neuronal activity using intrinsic signals. The CBF techniques rely mostly on tagging the blood spins differentially inside and outside of a well-defined volume. CBV techniques in the human exploit the fact that blood spin-lattice relaxation rate $T_1$ is different for tissue and blood so that the blood signals can be nulled while signals remain detectable from tissue and cerebral spinal fluid (CSF). Consequently, changes in CBV lead to a change in the total signal coming from an image voxel as the fraction of the voxel volume occupied by blood is altered in response to neuronal activity. In animal models, a commonly employed approach uses intravascular exogenous contrast agents that stay within the blood vessels for a long period of time (e.g. Mandeville et al. 1996; Mandeville et al. 1998; Mandeville and Marota 1999; Mandeville et al. 2001; Leite et al. 2002; Mandeville et al. 2004; Zhao et al. 2006; Goense et al. 2007; Smirnakis et al. 2007; Zhao et al. 2007), but this approach is not usable in humans because of toxicity issues. However, none of these approaches has been able to replace BOLD-based fMRI in human studies, because of a variety of shortcomings associated with the different techniques, such as smaller functional contrast-to-noise ratio, difficulty of rapid whole brain coverage, and/or inapplicability to humans because of toxicity issues (in cases of exogenous contrast).
MECHANISTIC CONSIDERATIONS

In the quest to improve spatial accuracy of functional images, efforts to understand the complex set of mechanisms underlying the coupling between deoxyhemoglobin containing blood vessels and MRI signals have been crucial. Modeling studies have been at the heart of this effort, starting with the first modeling paper from our group (Ogawa et al. 1993) immediately following the introduction of fMRI. Since then, several groups have published on this topic (e.g. Ogawa et al. 1993; Kennan et al. 1994; Weisskoff et al. 1994; Yablonskiy and Haacke 1994; Bandettini and Wong 1995; Boxerman et al. 1995a; Boxerman et al. 1995b; van Zijl et al. 1998; Kiselev and Posse 1999; Fujita 2001; Kjolby et al. 2006; Uludag et al. 2009). Here, some of the important aspects of these findings are summarized, without going into detail.

BOLD contrast reports on the overall deoxyhemoglobin content in an imaging voxel. In the brain, this is determined by two parameters: 1) the deoxyhemoglobin concentration in blood and 2) the total amount of deoxyhemoglobin containing blood volume in a given volume of brain tissue. The former is directly proportional to the CMRO₂/CBF ratio, i.e. the rate at which oxygen is consumed compared to its rate of delivery. Magnetic field inhomogeneities generated in and around blood vessels because of the presence of the paramagnetic deoxyhemoglobin depends on vessel diameter. Consequently, blood vessels play a critical role in the coupling between alterations in neuronal activity and MR detectable signals. The nature of the MR detectable signals also depends on parameters other than physiology, namely magnetic field magnitude and the type of pulse sequence used.

EXTRAVASCULAR EFFECTS

The BOLD effect has two components: extravascular and intravascular (i.e. blood related). First, let us focus exclusively on the extravascular space. When a sample, such as the human brain, is introduced into a static magnetic field, as encountered inside the magnet of an MRI instrument, the magnetic field in the sample is not necessarily equal to the value the magnet generates in air. That is because all matter, including biological tissue, has magnetic properties; consequently, an external magnetic field induces additional magnetic fields in any sample, which are generated by the sample itself in response to an external magnetic field. This is conceptual-
ized by the property called magnetic susceptibility. Most biological tissues, including oxygenated blood, are weakly diamagnetic, that is they generate a very weak magnetic field that opposes the external magnetic field. However, deoxyhemoglobin is strongly paramagnetic, creating an additional magnetic field that is aligned with the external magnetic field. Thus, oxygenated blood and tissue have similar magnetic susceptibility, but in the presence of dHb, magnetic susceptibility of blood is significantly different inside versus outside the blood vessel containing dHb. In the presence of a large susceptibility difference between inside and outside of a blood vessel, the magnetic field inside the blood vessel becomes significantly different than outside, and the homogeneity of the magnetic field outside the blood vessel is perturbed in the vicinity of the luminal boundaries of the blood vessel; this extravascular perturbation is illustrated in Figure 2, approximating for simplicity, the blood vessel as an infinite cylinder with a different magnetic susceptibility relative to its surrounding; many salient features of the physics of this problem can be understood with this simplification (See Appendix 1).

![FIGURE 2](image)

FIGURE 2: Magnetic field around a cylinder oriented perpendicularly to the external static field containing 0%, 10%, and 30% deoxyhemoglobin. The susceptibility between inside and outside the cylinder was taken to be the same at 0% deoxyhemoglobin. The field magnitude is represented as a three dimensional plot. (Calculations and the figure generated by Ute, Goerke, CMRR).

The magnetic field is homogeneous inside such a cylinder, and inhomogeneous outside in the vicinity of the cylinder. At any point outside the cylinder, the magnetic field will vary depending on the distance from the blood vessel and orientation relative to the external magnetic field direction. The magnitude of the magnetic field change induced outside the cylinder will be proportional to the applied external magnetic field, i.e. the magnetic field of the
instrument used (1.5 Tesla, or 7 Tesla etc.). This is fundamentally the origin of the relevance of field strength in fMRI and why we try to use higher magnetic fields such as 7 Tesla, instead of operating at 1.5 Tesla. There are, of course, many other considerations that come into the field dependence issue.

Outside the cylinder, the magnetic field is directly proportional to \((r_b/ r)^2\) where \(r_b\) designates the cylinder radius, \(r\) is the distance from the point of interest to the center of the cylinder in the plane normal to the cylinder. Thus, the magnetic field perturbation induced by the susceptibility difference changes rapidly over a distance comparable to two or three times the cylinder radius; at a distance equal to the diameter of the cylinder from the cylinder center, the field is already down to 25% of its value at the cylinder boundary and close to what it would be if the blood vessel was not there at all.

What is of critical importance for the extravascular BOLD effect is that the magnetic field is inhomogeneous around dHb containing blood vessels. The magnitude of this inhomogeneity is proportional to the external magnetic field applied, and the magnetic susceptibility difference between inside and outside the blood vessel, which in turn depends on the amount of dHb present; the spatial extent of the inhomogeneity is about a diameter or so around the blood vessel. This implies that the resonance or precession frequency\(^3\) of spins varies over space in the voxel where there exists dHb containing blood vessels, i.e. capillaries, venules and veins.

In a GRE fMRI experiment, images are acquired after a delay TE in order to sensitize the image to magnetic field inhomogeneities. Intra-voxel magnetic field inhomogeneities lead to signal loss during the evolution period TE, because the precession frequencies of the spins are non-uniform within the voxel; magnetization vector of the voxel, which is the vector sum of all the magnetic moments associated with each water proton, will decrease with time as the different populations of these protons, encountering different magnetic fields, will have different precession frequencies. Therefore, they will not stay in phase. If there are many such vessels with different orientations and dimensions, this dephasing can be expressed as a random process obeying an exponential decay. In general, the time constant \(T_2^*\) characterizes the exponential decay of MR signals after they are excited. Thus, presence of dephasing due to dHb-containing blood vessels in a voxel contributes to the exponential decay and leads to a shorter \(T_2^*\). If the dHb content is altered, the
magnetic field gradients will also be altered (Figure 2), and consequently, the signal loss due to dephasing and the time constant $T_2^*$ will also change.

Diffusion introduces a modification to the mechanism described above. The diffusion distances of water in tissue during the typical TE’s (~20-50 ms) feasible in an fMRI experiment is such that the water molecule can sample all the magnetic field variations around small blood vessels, such as capillaries or small venules that immediately follow capillaries (e.g. Ogawa et al. 1993; Yablonskiy and Haacke 1994; Boxerman et al. 1995a; Boxerman et al. 1995b; Kennan et al. 1998; Kiselev and Posse 1999; Fujita 2001; Uludag et al. 2009). This random motion averages the magnetic field gradients around these small vessels, resulting in a substantially diminished BOLD effect. The averaging does not occur around larger blood vessels because the magnetic field gradients occur over longer distances, dictated by the blood vessel diameter. The diffusion mediated BOLD effect associated with the microvasculature is small but it increases supralinearly with magnetic fields up to ~7T (Uludag et al. 2009).

An alternative but rarely used approach for functional mapping utilizes spin-echoes (SE) where a single refocusing pulse, typically a 180° pulse, is inserted in the delay period TE between signal excitation and acquisition. The refocusing pulse leads to a rephasing of the dephasing that was induced by the magnetic field inhomogeneities within the voxel, thus recovering the signal loss that occurs due to dephasing. But the diffusion effects are not recovered since the path of the diffusing spins is not the same on the two sides of the refocusing pulse because diffusion is a random motion. Thus, in a SE fMRI experiment, the contribution of blood vessels beyond the microvasculature (e.g. with diameter $\geq 10\, \mu m$) to functional signals are attenuated significantly; instead, the microvascular contributions dominate. However, this diffusion-mediated effect associated with the microvasculature is small, as previously stated, accounting for the unpopularity of SE fMRI; but it increases supralinearly with magnetic field up to ~7T (Uludag et al. 2009) reaching robust and easily detectable levels at that field strength (Yacoub et al. 2003; Yacoub et al. 2005) that are, in fact, somewhat comparable to what one observes at 3T.

In a SE measurement, the amplitude of the spin-echo detected decays exponentially with a time constant $T_2$. If the magnetic field within a voxel in the sample is completely homogeneous, then, for
that voxel $T_2 = T_2^*$. But even in the absence of deoxyhemoglobin this condition is rarely fulfilled. So, $T_2^* \leq T_2$ always. Signal decay in a GRE experiment, characterized by the parameter $T_2^*$ also contains the contribution of processes that contribute to $T_2$. In fact, under most circumstances, one can simply write $1/T_2^* = 1/T_2 + 1/T_2'$ where the second term accounts for all other effects besides $T_2$.

Diffusion mediated change in a SE amplitude in the presence of magnetic field inhomogeneities is actually not a true $T_2$ effect because the effective $T_2$ associated with this process depends on the TE itself. For very short TE’s, diffusion induced motion of spins becomes small and recoverable under a refocusing pulse; this is not the case for longer TE values. In contrast, processes like dipole-dipole interactions between spins result in signal loss in a SE experiment that can be characterized with a constant $T_2$ that is independent of the echo time TE. Nevertheless, in the fMRI world, people talk about $T_2$ effects when referring to functional signal changes detected in a SE experiment. But we should really talk about an apparent $T_2$ to remain cognizant of the fact that $T_2$ is not just a constant independent of TE for the BOLD effect.

INTRAVASCULAR EFFECTS

In blood, hemoglobin is also compartmentalized within red blood cells. Therefore, when the deoxy form is present, there are field gradients around the red blood cells. However, because the dimensions of these gradients are very small compared to diffusion distances, the effect is averaged by diffusion, as it is for capillaries and small venules, and becomes a change in the apparent $T_2$ only. This dynamic averaging also involves exchange across the red blood cell membrane which is highly permeable to water. Thus, the apparent $T_2$ of blood changes when the content of deoxyhemoglobin is altered by elevated neuronal activity and this leads to a signal change in a $T_2$ weighted image. This effect also comes into $T_2^*$ weighted image because, as previously stated, all processes that contribute to $T_2$ also contribute to $T_2^*$.

This change in blood apparent $T_2$ (and consequently a signal intensity change in a $T_2$ and $T_2^*$ weighted image) will be present wherever and whenever the content of deoxyhemoglobin has changed; thus, it will be associated with both large and small blood vessels. This effect is substantial, even the dominant contribution to functional imaging signals at low magnetic fields such as 1.5 Tesla (van Zijl et al. 1998; Duong et al. 2003). It diminishes in importance
at high fields because the $T_2^*$ and $T_2$ of blood decreases precipitously with increasing magnetic field magnitude and becomes much less than that of signals associated with tissue (Duong et al. 2003; Uludag et al. 2009); consequently, in an acquisition where the TE is set approximately equal to tissue $T_2^*$ or $T_2$ for GRE and SE fMRI, respectively, blood signals are significantly attenuated and are even diminished beyond detectability. At 3 Tesla, blood signals were shown to account for approximately 50% of the SE based fMRI signals. At 7 Tesla, the upper limit on these contributions is less than about 10% at TE values corresponding to tissue $T_2$ (Duong et al. 2003).

**INFLOW EFFECTS IN BOLD-BASED FMRI**

During the hemodynamic response that supplies more blood to the area of increased activity in the brain, flow increases must also take place within the vascular tree supplying and draining the activated area. This vascular flow change can lead to signal alterations in large blood vessels in images intended to report on BOLD contrast (e.g. Menon et al. 1993; Duyn et al. 1994; Frahm et al. 1994; Kim et al. 1994; Segebarth et al. 1994). This is not because BOLD contrast itself contains a direct inflow effect; rather, image contrast may not be purely of BOLD origin because any repeated, slice-selective image is inherently flow sensitive if the signal within the slice does not attain full relaxation between consecutive signal excitations. In single slice studies, allowing full relaxation in between RF pulses eliminates this problem completely; however, this condition often is not satisfied in many studies since it leads to a loss in SNR per unit time. Consequently, such studies essentially obtain images of macrovascular flow rather than the much smaller BOLD changes. This problem was demonstrated with clarity by comparing presumably BOLD-based functional images with vessel images in two and three dimensions (Segebarth et al. 1994; Belle et al. 1995).

In a multi-slice experiment, even if the time between the acquisition of the same slice in an fMRI time series is compatible with full relaxation of spins in that slice, an inflow problem may still be present due to inter-slice effects. For example, blood experiencing an RF pulse in one particular slice will subsequently travel to other parts of the brain and, in turn, affect the intensity of signals from a slice sampled at a later time. The sensitivity to this problem will depend on the imaging sequence used, and how the different slices are sampled and oriented, since most macrovascular flow occurs along
an inferior-superior direction in the human brain. Experimental evidence, however, indicates that the contribution of this type of inter-slice flow to BOLD-based multi-slice imaging is negligible (Howseman et al. 1999).

MODELING OF FMRI SIGNALS

A complete modeling of stimulus- or task- induced fMRI signal changes must consider the aforedescribed extravascular and intravascular mechanisms as well as the numerous basal physiological parameters, such as blood volume, and blood oxygenation, and the changes induced in these parameters by alterations in brain activity. Such a comprehensive model was recently published (Uludag et al. 2009). This paper also presents a good review of what has been done previously in this area, and the interested reader is referred to it for an overview of all modeling efforts and relevant literature.

Figure 3 illustrates the calculated fractional signal change for a spin-echo fMRI experiment for macro and microvasculature from the Uludag et al. (2009) paper. The macrovascular plot presented was calculated for a blood vessel orthogonally oriented to the main (static) magnetic field, and thus represents an upper limit for the macrovascular contribution; this contribution would decreases precipitously when the orientation of the blood vessel deviates from orthogonal, and is absent when the blood vessel is completely and perfectly parallel to the external magnetic field in the voxel. For the microvasculature, the orientation does not come into play because there are so many such vessels of different orientations in a typical MRI voxel; so, an average of all orientations is taken.

Figure 3 demonstrates the clear dominance of microvascular effects in a SE experiment and that this dominance increases with increasing magnetic field. There is some uncertainty as to what the microvascular contribution is to the SE fMRI signal; two different plots are given for it and they correspond to a condition where CBV changes accompanying neuronal activity alterations do or do not occur in the dHb containing microvasculature, the only CBV change that is relevant for the SE BOLD effect. While there are good estimates of CBV changes correlated with alterations in CBF, and hence with stimulus evoked activity (e.g. Grubb et al. 1974; Grubb et al. 1978; Ito et al. 2001), the measurements pertain to the total CBV associated with the entire vascular tree, from arteries to draining veins. Assuming this total fractional CBV change is applicable to all vessels, including the dHb-containing microvascula-
ture, this represents one limit for the modeling results of SE fMRI presented in Figure 3 with the solid black line. It is likely, however, that CBV changes do not occur, or do not occur to the same extent, in all parts of the vasculature. In fact MR studies suggested that dHb containing blood vessels from capillaries to draining veins do not undergo CBV changes coupled to neuronal activity (Kim and Kim 2005; Kim et al. 2006) at all, in agreement with PET data that takes into account the different CBV components in in the modeling of radiotracer kinetics (Ito et al. 2005). This represents the other limit (i.e. no CBV change coupled to neuronal activity alteration in the microvasculature) illustrated with the dashed line in Figure 3. For GRE fMRI, large vessels always contribute to the fMRI mapping signals and always yield the largest effects. They cannot be avoided in GRE fMRI at any field strength. However, as the field magnitude increases, the relative contribution of the microvasculature increases as well. This is illustrated for the extravascular effect in Figure 4.

FIGURE 3: Simulation results for SE fMRI for TE= T₂ as a function of field strength. Total fMRI signal (considering intravascular and extravascular effects) is shown in black for microvasculature assuming no stimulus induced CBV change (dashed line), and 16% microvascular CBV change. Most probable scenario is between these two extremes. Total fMRI signal for macrovasculature is shown in magenta for the extreme case of a vessel oriented perpendicularly to the static magnetic field, which causes the maximum effect. Microvascular CBV was taken to be 2.5% and was assumed to be composed of 20% arterioles (diameter = 16 µ), 40% capillaries (diameter = 5 µ), and 40% venules (diameter = 16 µ). Adapted from Uludag et al. 2009.
FIGURE 4: Relative contribution to $T_2^*$ weighted GRE fMRI from capillaries vs. larger blood vessels (veins) as a function of the frequency on the vessel wall for a cylinder mimicking blood vessels. There are two curves displayed, in red and blue, representing the ratio of contributions from a 5 µ blood vessel relative to a 60 µ and to a 100 µ vessel, respectively. The two are virtually superimposable, since the contribution becomes independent of vessel diameter after about ~10 µ. The ratio is calculated for the same susceptibility difference between vessel interior and the surrounding tissue for both the 5 µ vessel and the 60 and 100 µ vessels. In vivo, this is not the case because a 5 µ vessel, which is a capillary, has a lower average deoxyhemoglobin content than the veins draining the capillary bed. Taking this into account would not change the overall picture but would shift this curve to the right. Note that the horizontal axis is dependent on the product of the deoxyhemoglobin content and the external magnetic field applied. Fixing the deoxyhemoglobin content at 40% yields the frequency on the vessel wall corresponding to magnetic fields of 1.5, 7 and 10.5 Tesla, indicated by the vertical dashed lines. Adapted from Uludag et al. 2009.

SPATIAL SPECIFICITY OF BOLD FMRI

Although the neuro- and cognitive- science communities have embraced fMRI with exuberance, caution must be exercised in quantitatively using fMRI data because fMRI maps are based on secondary metabolic and hemodynamic events that follow neuronal activity, and not the electrical activity or the neurotransmission directly. One of the most important issues is the spatial specificity of the fMRI maps, i.e. how accurate are the maps generated by fMRI compared to actual sites of neuronal activity (e.g., Ugurbil et al. 2003a; Ugurbil et al. 2003b).
Neuronal organizations that respond to a highly reduced attribute of an input and perform elementary computations are known to exist in the millimeter to submillimeter spatial scale both tangentially on the cortical surface and across the cortical laminae. Columnar clustering is likely to be a fundamental organizational principle in the brain. Based on monkey electrophysiology studies, clustering analogous to the well-recognized columns in the early visual areas has been proposed for higher order visual areas such as the middle temporal (MT) for direction of motion (Albright 1984; Albright et al. 1984), and inferio-temporal (IT) cortex for face and object recognition (Fujita et al. 1992; Wang et al. 1996). Imaging such organizations, however, requires submillimeter specificity of mapping signals and sufficiently high image SNR to meet the spatial resolution demands in image acquisition. This is an important technical challenge in fMRI. This challenge is illustrated by the fact that the point spread function (PSF) of GRE fMRI, the most commonly used form of fMRI, was measured to ~3.5 mm at 1.5 and 3 Tesla (Engel et al. 1997; Parkes et al. 2005). When areas of the brain containing draining veins were avoided, the PSF upper limit was measured to be 2 mm at 7 Tesla (Shmuel et al. 2007). These numbers are significantly larger than the spatial scale of cortical columns or the different neuronal layers across the cortical thickness.

In the two decades since the introduction of fMRI, functional mapping with columnar and laminar resolution seems to have been largely uninteresting in the human cognition community. In part, this must be due to the difficulty of reaching this level of resolution with any form of MR imaging because of signal-to-noise limitations. The fact that most cognitive studies performed on the human brain have not yet reached a mechanistic formulation at this elementary level must have also played a role. Nevertheless, at CMRR we decided to pursue the question of columnar or laminar functional mapping immediately after the introduction of fMRI. Particularly influential for us was the work by our colleague Apostolos Georgopoulos using electrophysiology in non-human primates. In a series of ground-breaking studies published in Science (Georgopoulos et al. 1986; Georgopoulos and Grillner 1989; Georgopoulos et al. 1989; Georgopoulos et al. 1992; Georgopoulos et al. 1993), Georgopoulos had demonstrated the presence of directionally tuned cells in the primary motor cortex and introduced the concept of the population vector (Georgopoulos et al. 1986); when individual, directionally tuned, cells were represented as vectors according to changes in
their activity during a movement of the arm, the construct that is
the vector sum of all these cell vectors (population vector) was in a
direction congruent with the direction of movement. This is a for-
ward, predictive, and mechanistic model linking neuronal activity
with behavior. It has become relatively common in electrophysi-
ology experiments. To construct such forward models with fMRI
at this elementary level, however, most likely requires relatively
high resolution (at or approaching columnar resolution) and spatial-
ly specific (to sites of neuronal activity) imaging that can provide
numerous functionally distinct and independent voxels. Achieving
this type of capability provided the motivation for the steady effort
at CMRR towards attaining improved accuracy of functional im-
ageing signals. This effort is largely responsible for the plethora
of animal model studies exploring the fidelity of functional signals to
sites of altered neuronal activity (e.g. Duong et al. 2000; Duong et
al. 2001; Harel et al. 2002; Harel et al. 2006; Lee et al. 1999; Silva
et al. 2000; Lee et al. 2001; Lee et al. 2002; Kayser et al. 2004; Kim
et al. 2004; Zhao et al. 2006) and development of ultrahigh field
magnetic resonance imaging (e.g. Ugurbil et al. 1999; Ugurbil et al.
2000; Ugurbil et al. 2003a). In this effort, I regard one of our papers
(Duong et al. 2001) as critically important since it demonstrated
for the first time that blood flow is regulated at the level of orienta-
tion columns in the cat visual cortex (discussed further in the next
section). Parallel to the animal model studies, the development of
7 Tesla for human MRI, which was first accomplished in CMRR,
was also largely propelled with the drive to attain high-resolution
functional mapping with high spatial fidelity to sites of neuronal
activity.

Ultimately, these efforts led to robust imaging of ocular domi-
nance columns at 7 Tesla (Yacoub et al. 2007) (after earlier results
at 4 Tesla by us (Menon et al. 1997) and others (Cheng et al. 2001;
Goodyear and Menon 2001)), the first time imaging of orientation
columns in the human brain (Yacoub et al. 2008), and the axis of
motion selective features in human area MT (Zimmermann et al.
2011). The MT work also demonstrated that these organizations
could be mapped with laminar resolution, supporting our other 7
Tesla work on imaging cognitive processing at different layers of
the cortex in V1 (Olman et al. 2012). Figure 5 illustrates the func-
tional images obtained in the human MT on either three flattened
layers at different depths or in planes that run perpendicular to the
cortical surface.
FIGURE 5: Axis of motion columnar mapping in one subject. (a) Results of MT localization are projected onto the cortical reconstruction of the subject's left hemisphere (neurological convention). High-resolution cortical grid sampling is performed. Zoomed-in view to the subject's STS shows overlaid streamlines at three relative cortical depths. (b) Results of the high resolution cortical grid sampling for the motion direction experiment showing columnar organization of axis of motion features in three sampled layers. Top layer may be degraded by residual contribution from pial vessels. (c) Representation of the high resolution cortical grid sampling showing four three-dimensional vertical slices through the sampled cortical layers depicting the consistency of cortical columns tangential to the surface. Adapted from Zimmermann et al. 2011.

A previously unknown columnar organization that responds preferentially to either low or high temporal frequency was imaged recently at 4T (Sun et al. 2007). In collaboration with CMRR, Georgopoulos also demonstrated an fMRI based population vector construct in the human parietal cortex with 4 and subsequently 7 Tesla fMRI, predicting behavior in a maze task which had a direction component embedded in it (Gourtzelidis et al. 2005; Jerde et al. 2008) (Figure 6); in these studies, the fMRI data were shown to provided virtually identical results to electrophysiology-based measurements in the trained nonhuman primates (Figure 6) that were performed independently (Crowe et al. 2004). It was also demonstrated in these experiments that 7T provided a significantly larger number of independent voxels that were identified as being unique-
ly tuned (Jerde et al. 2008). These are the first fMRI examples of encoding modeling that are currently generating significant interest (Naselaris et al. 2011).

FIGURE 6: Upper figure (left): Directional tuning curve for voxels exhibiting tuning for 270 degree maze exit. Upper figure (right): Standardized population tuning curve (see Gourtzelidis et al., 2005) for voxels and cells. The cell tuning curve is from Fig. 6 in (Crowe et al., 2004) and was obtained with electrophysiological recordings in trained non-human primates. Lower figure: The direction of the population vector computed from all the data from the human experiments plotted against the corresponding direction of the maze path. Adapted from Gourtzelidis et al. 2005.

However, I regard these accomplishments as a prelude to a new era where whole brain high-resolution fMRI, at or approaching columnar and layer resolution is feasible at 7T. This capability has been used recently to perform 1 mm isotropic resolution whole brain resting state fMRI (De Martino et al. 2011). Figure 7 illustrates 0.75 mm isotropic whole brain single-shot EPI images acquired at 7T; these are excellent quality EPI images, better than even what most people use at much lower field strengths where EPI is easier to execute. They also have sufficient SNR to serve in an fMRI study. I believe 0.5 mm isotropic will be attainable when 7 Tesla systems with more than 32 receive channels will become available. Numerous technological and engineering solutions had
FIGURE 7: Whole brain EPI images at 7 Tesla: 0.75 mm isotropic Echo spacing 0.67 ms, 256 x 256, parallel imaging with reduction factor 4. Partial Fourier = 6/8 TE = 20 msec, 128 slices TR = 6 s, single image. Data generated in CMRR, University of Minnesota by Yacoub et al.

to be realized at 7 Tesla before such imaging capability became feasible. These include parallel imaging to accelerate along the phase encoding direction (Pruessmann et al. 1999; Sodickson et al. 1999; Griswold et al. 2002) (probably the most important development for ultrahigh field imaging using single-shot EPI, which is the workhorse of fMRI), slice accelerated whole brain functional imaging (Moeller et al. 2008; Moeller et al. 2010; Feinberg et al. 2010), and parallel multichannel transmission (e.g. Adriany et al. 2005; Van de Moortele et al. 2005; Vaughan et al. 2006; Metzger et al. 2008; Setsompop et al. 2008; Setsompop et al. 2008; Setsompop et al. 2008; Setsompop et al. 2009); in fMRI, accompanying anatomical images are also of great importance. Yet at 7 Tesla, obtaining good anatomical images without large intensity variations that obscure tissue contrast and/or preclude automated gray-white matter segmentation was initially difficult. This problem was solved with the
Two Decades of fMRI

introduction of biased field correction methods in anatomical imaging (e.g. Duyn et al. 2007; Van de Moortele et al. 2009). Coupled with improved anatomical contrast (e.g. Duyn et al. 2007; Rooney et al. 2007; Budde et al. 2011; Henry et al. 2011) now available, ultrahigh fields provide superb anatomical images that can accompany functional images at these field strengths.

SPECIFICITY OF PERFUSION CHANGES COUPLED TO NEURONAL ACTIVITY

In fMRI, the degradation of spatial specificity might arise from imprecise spatial-coupling between neuronal activity and the physiological and metabolic events that ultimately yield the functional maps. The dominant physiological change induced by alteration in the neuronal activity relevant for fMRI is CBF change. There is no a priori requirement that control of CBF spatially match functional organizations as long as CBF change includes the functional territory responsible for invoking the altered CBF state. In fact, at the time of the introduction of fMRI, prevailing concepts suggested that CBF control far exceeds the territorial boundaries of columnar organizations; in other words, “the brain watered the entire garden for the sake of a single thirsty flower” (Malonek and Grinvald 1996). We examined this question again using newly developed MR methods (Duong et al. 2001) and showed that in fact, the brain waters the thirsty flower while it sprinkles an extensive region around it.

Images of blood flow and blood flow changes associated with increased neuronal activity can be obtained using arterial spin labeling (ASL) techniques that utilize the water protons in the blood as an endogenous transient tag. These methods rely on either continuous (e.g. Detre et al. 1992; Zhang et al. 1993; Detre et al. 1994) or dynamic (i.e. modulated versions of continuous (Barbier et al. 1999; Barbier et al. 2001) or pulsed (e.g. Edelman et al. 1994; Kim 1995; Wong et al. 1998) tagging approaches. Efficiency of some of these methods was recently evaluated quantitatively (Pohmann et al. 2010). These ASL techniques can be tuned to be selectively sensitive to capillary/tissue level flow, and the detection of blood flow or blood flow changes in large blood vessels can be minimized. Such maps report on water delivery to the capillary bed and, by exchange across the capillary wall, to surrounding tissue, i.e. perfusion.

In these MR techniques, a perturbation (the label) is induced in the population of hydrogen nuclei (spins) of water in blood, outside the region of interest; this label is monitored as it shows up in the
tissue in the region of interest after a delay (the tagging time) that is long enough to permit arrival into the capillaries and the surrounding tissue but not long enough to reach thermal equilibrium. In this case, the amount of label detected in the tissue in the region of interest is proportional to blood flow, and increases with elevated neuronal activity. Since the arterial side is permeated with fresh blood significantly faster relative to the tissue, the time to reach the equilibrium state in the large arteries versus the tissue is different. Thus, long tagging times (e.g. ~1.5 to ~2 s) (Tsekos et al. 1998) eliminate the arterial component from perfusion images. The long tagging times, however, may lead to the tagged spins appearing in the venous side, thus leading to false activation in the draining veins. At high magnetic fields, the $T_2$ and $T_2^*$ of venous blood is very short (Thulborn et al. 1982; Lee et al. 1999) so that this effect can be selectively eliminated by a brief delay after excitations of spins but before image acquisition. Accordingly, perfusion-based fMRI maps have been shown to yield accurate images that co-localize with Mn$^{+2}$ uptake (Duong et al. 2000), a marker of calcium dependent synaptic activity (Lin and Koretsky 1997).

When potential large vessel contributions are suppressed with appropriate choice of imaging parameters, perfusion-based fMRI can be used to examine the critical physiological question related to specificity of blood flow increases. A specific question that can be asked is if perfusion changes are confined accurately to the region of increased neuronal activity in the spatial scale of columnar organizations?

Such a study was performed using the iso-orientation columns in the cat visual system (Duong et al. 2001) using single condition mapping (stimulus vs. a non-stimulus control state); the results demonstrated that while perfusion increases that follow neuronal activation were not perfectly localized at the iso-orientation column level, the difference between active and neighboring inactive columns was large and permitted single condition mapping (Figure 8). In other words, as previously stated, the brain was found to literally water the thirsty flower while sprinkling an extensive region around it. To accomplish this, there must exist regulation of blood flow control at the submillimeter (~300 to 400 µm) scale. These data were also used to calculate a point-spread function for the CBF control in the cat cortex; this PSF, at full width at half-maximum (FWHM), was 470 µm under single-stimulus condition without differential subtraction.
FIGURE 8: Activation maps of two orthogonal iso-orientation domains in the cat visual cortex obtained separately with perfusion based functional imaging mapping. Panel a and b, show perfusion based images obtained with two orthogonal orientations (45 and 135 degree gratings), demonstrating their complementarity. Each map was acquired as a single condition map where the grating of one orientation was either moving back and forth (activation condition) or static (control condition). Bottom trace (Panel c) shows the cerebral blood flow (CBF) response at the tissue level (i.e. perfusion) in all the voxels identified as activated either by the 45 degree or 135 degree grating for the two orthogonal stimulations. All voxels identified as activated for 45 degree orientation display a large perfusion increase during stimulation by this orientation gratings and show a smaller but detectable perfusion increase in response to the 135 degree gratings. The converse is observed for the voxels identified as activated for 135° gratings. A marked perfusion increase (~55-60 %) following 45° or 135° stimulus was observed in the regions tuned to these orientations, while the stimulation with the orthogonal orientation lead to a 3.3± 0.6 fold smaller perfusion increase in the same region. Adapted from reference (Duong et al. 2001).

It is important to emphasize that the perfusion based images of the orientation columns in the cat cortex (Duong et al. 2001) described above was obtained under single condition as opposed to differential mapping. Differential mapping refers to functional images obtained by effectively subtracting two analogous but orthogonal activation states; this approach is assumed to generate
functional maps with high fidelity to the territory of altered neuronal activity because spatially non-specific mapping signals are assumed to be common to the orthogonal activation conditions. This assumption appears to be incorrect at the level of cortical columns (Shmuel et al. 2010). Nevertheless, differential mapping is expected to have better spatial accuracy than signal condition maps, which do not employ any strategy to suppress spatially inaccurate mapping signals. Mapping ocular dominance columns by stimulating one eye vs. the other in an alternating fashion is differential mapping. Stimulating one eye and using a dark state as the control would correspond to single condition mapping. Phase encoded mapping using an activation paradigm that cycles through several different possible stimulations, as employed in the visual system (e.g. Sereno et al. 1995; DeYoe et al. 1996; Engel et al. 1997), is also a kind of differential mapping because it would suppress common, non-modulating components.

Thus, the ability to obtain single-condition maps of orientation domains by perfusion imaging (Duong et al. 2001) was a fundamentally important result for brain physiology and functional mapping with magnetic resonance because it demonstrated for the first time that CBF changes are regulated even at the level of capillaries in addition to being subject to upstream control at the level of arterioles and arteries. This conclusion was initially considered to be in conflict with results from optical imaging techniques that report on CBV, which is assumed to be correlated with CBF in space and magnitude. However, there is an important difference between optical imaging and MR perfusion imaging data. Namely, the MR methods can be set to eliminate, in the image acquisition phase, the confounding problem coming from large vessels. When the origin of the functional mapping signals is restricted to the spatially accurate capillary/tissue level by the physics of the MR acquisition method, accurate columnar activation is detected. However, the optical methods do not have this selective tuning capability and report on all CBV changes; thus, they can in fact be dominated by large vessel effects. Recent optical imaging studies conducted subsequent to the perfusion mapping of iso-orientation columns by perfusion fMRI, have confirmed this fundamental difference and demonstrated that when large vessel effects are taken out, CBV based optical imaging techniques also yield columnar level mapping signals (e.g. Harrison et al. 2002; Sheth et al. 2003; Vanzetta et al. 2004). Subsequently, CBV based fMRI studies, conducted using the extravascular con-
contrast agent MION, have also demonstrated single condition columnar resolution (Zhao et al. 2005; Harel et al. 2006); these MION based fMRI studies, with their superior sensitivity, further revealed that laminar activity can be spatially resolved and that the larger intralaminar signal changes correspond to cortical layer 4 (Harel et al. 2006). These observations further confirm that capillary level control of blood flow and volume in the brain exists and modulates activation-induced alterations in these physiological parameters at the columnar and laminar level.

**SPATIAL ACCURACY OF MAGNETIC RESONANCE SIGNALS IN FUNCTIONAL MAPPING**

Although the spatial fidelity of coupling between neuronal activity and the physiological and metabolic events that ultimately yield the functional images contain sufficient precision to map columnar structures, there is another potential source of degradation of spatial accuracy, which has been already alluded to; namely, the coupling between the MRI signals and the physiologic changes that occur. This coupling in turn depends strongly on magnetic field and the pulse sequence used. Many of the issues that exist with respect to this question were already mentioned in discussing the BOLD mechanism. They are revisited and summarized here, organized in a more concise form.

All vessel types containing dHb, ranging from capillaries to large veins contributed to the $T_2^*$ GRE fMRI, the overwhelmingly dominant techniques used in fMRI. However, as previously mentioned, the large vein contribution dominates GRE fMRI at all magnetic field strengths. Thus, GRE fMRI suffers from inaccuracies in functional mapping because of large vessel contributions. Often one is concerned with the large blood vessels on the pial surface of the brain. Deoxyhemoglobin changes clearly must occur first at the location where stimulus or task induced neuronal activity changes take place; however, due to blood flow, these dHb changes must propagate down the vascular tree to draining veins on the pial surface. At some point in the vascular tree, the functionally induced changes will be diluted by blood that is pooled from other unaffected regions by the task or the stimulus, and hence will disappear from the functional map. But, before this dilution occurs, functional signals associated with draining veins will be detected; that this represents a major confound was realized and documented as early as in 1993 (Menon et al. 1993; Kim et al. 1994) and was
recently further reemphasized (Polimeni et al. 2010). GRE fMRI studies would also have a strong contribution from \(~50\) to \(~100\ \mu\) intracortical veins that run perpendicular to the cortical surface and drain the different layers. These are the veins that were visualized in the original BOLD imaging work conducted in rats (Ogawa and Lee 1990); they can also be seen in very high-resolution human brain images with high \(T_2^*\) or phase contrast (e.g. Vaughan et al. 2006; Budde et al. 2011).

The GRE BOLD effect comes from both extravascular and intravascular (blood) sources. The blood contribution to fMRI signals can be the dominant source of functional maps at lower fields. This blood effect occurs at all levels of the vasculature. Its presence in capillaries and small venules is not a problem for spatial specificity; however, the functional signals associated with blood in large veins degrade the fidelity between regions of altered neuronal activity and the fMRI maps. This confound is diminished at ultrahigh fields as blood \(T_2^*\) decreases (Duong et al. 2003; Uludag et al. 2009) and becomes much shorter than tissue values of this parameter. Consequently, in a typical fMRI acquisition where echo time \(TE\) is set approximately equal to tissue \(T_2^*\), blood contribution to the image can be significantly reduced or even absent with increasing magnetic field strength beyond 3 Tesla (Duong et al. 2003). Hence, at ultrahigh fields such as 7 Tesla or higher, the intravascular blood effect, one of the major sources of inaccurate functional signals, is suppressed or even eliminated.

The extravascular BOLD effect for large vessels persists at all field strengths and even increases with increasing field magnitude (Uludag et al. 2009); hence, it continues to be a source of inaccurate functional mapping signals in GRE fMRI. However, the microvascular contributions also get significantly larger at the ultrahigh fields and become comparable to the large vessel effects (Yacoub et al. 2001; Yacoub et al. 2005), enabling accurate mapping that would be difficult to achieve at lower magnetic fields. This relative microvascular gain was demonstrated in a 4T-7T comparison (Yacoub et al. 2001) and was utilized early in the history of 7T fMRI studies to reveal the tonotopic organization of the human auditory cortex for the first time (Formisano et al. 2003), at a time when attempts at studying this organization at lower fields had failed to produce convincing data. Similarly, fine-scale single-digit activations in subdivisions of the human primary somatosensory cortex (SI) in individual subjects was accomplished recently using GRE fMRI at 7T.
(Stringer et al. 2011). Recently, the presence of this microvascular contribution at 7T was relied upon when it was demonstrated that in high resolution 7T fMRI studies, stripping away the outer cortical layers where the pial vein contribution is dominant and looking into deeper layers of the cortex (Polimeni et al. 2010) yielded more accurate functional maps; without the presence of the microvascular contribution independent of and in addition to the large vessel contribution on the surface, this strategy would not have worked with GRE fMRI.

The spatial inaccuracies introduced in functional mapping due to the large vessel confound present in GRE fMRI have been quantified experimentally in several ways. It was rigorously demonstrated that the GRE fMRI method fails to generate single-condition functional images of iso-orientation domains in the cat cortex (Duong et al. 2000; Kim et al. 2000) at 4.7 Tesla; the images of two orthogonal stimulation conditions were not complementary, as would be expected; rather, the highest activity was associated with a large draining vein, the sagittal sinus for the two orthogonal conditions. Simultaneous multiple site single unit recordings and fMRI studies on the same animal indicated that the limit of spatial specificity of GRE fMRI was in the 2 to 3 mm range for single-condition maps (Kim et al. 2004; Ugurbil et al. 2003b). At 1.5 and 3 Tesla human brain studies, the full width at half maximum (FWHM) of the point spread function (PSF) for GRE fMRI was estimated to be 3.5 mm (Engel et al. 1997) and 3.9 mm (Parkes et al. 2005), respectively. In contrast, at 7 Tesla, the PSF at FWHM was measured to be ~2 mm or less in the human brain in regions devoid of the large veins visible in the image (Shmuel et al. 2007), demonstrating advantages of the ultrahigh field strength.

The large vessel effects that persist in GRE fMRI even at ultrahigh fields can be suppressed with spin echo based fMRI at ultrahigh but not at low magnetic fields. As previously discussed, SE fMRI responds to apparent T2 changes both in the extravascular space around microvasculature (capillaries and small post-capillary venules) and in blood itself (van Zijl et al. 1998; Ugurbil et al. 1999; Ugurbil et al. 2000; Uludag et al. 2009). The spatially inaccurate intravascular effects in veins are suppressed at ultrahigh fields because, the apparent T2 of venous blood decreases precipitously with increasing magnetic field magnitude (see Duong et al. 2003; Uludag et al. 2009 and references therein); it is diminished from ~180 ms at 1.5 Tesla (Barth and Moser 1997) to ~6 ms at 9.4 Tesla (Lee et
Vol. 4  Cognitive Critique  151

al. 1999), significantly smaller than brain tissue T₂; thus, at TE values that correspond approximately to gray matter T₂, which would be optimal for detection of functional mapping signals associated with parenchyma at such field strengths, the blood signal would be significantly diminished, and even undetectable. At 3T, ~50% of the SE fMRI signals were shown to arise from blood (Norris et al. 2002). But at 7T, the blood contribution was estimated to be less than 10% at echo times matching gray matter T₂ (Duong et al. 2003). Consequently, at very high fields such as 7 Tesla or higher, the functionally non-specific blood component as well as the extra-vascular BOLD effect associated with veins are suppressed in SE studies (Duong et al. 2003; Yacoub et al. 2003; Uludag et al. 2009). Residual large vessel effects are still expected because the extravascular BOLD effect with large blood vessels is not exactly zero in SE fMRI; rather it is small and significantly less than effects associated with microvasculature (Uludag et al. 2009). This is the reason why SE fMRI techniques at ultrahigh fields yield robust columnar level mapping, providing maps that are uninterrupted by large vessel contributions (Yacoub et al. 2008).

HIGH RESOLUTION WHOLE BRAIN FMRI

The advantages of high and ultrahigh field fMRI have largely been exploited in applications where a subsection of the brain, such as the primary visual cortex (e.g. Yacoub et al. 2007; Yacoub et al. 2008; Olman et al. 2010) or the auditory cortex (e.g. Formisano et al. 2003) was studied using field of view restriction. Ultimately, however, we need to cover the entire brain in high resolution. Only recently, significant progress has been made in tackling the difficulties inherent in obtaining rapid whole brain coverage at 7 Tesla using techniques like EPI, the predominant method employed in fMRI. EPI is challenged by the shorter T₂* and increased magnetic field inhomogeneities at high fields. This difficulty is amplified with higher resolution as the EPI echo train length is elongated commensurately with resolution. Segmentation has been used in EPI to reduce the echo train length (McKinnon 1993; Feinberg and Oshio 1994; Wielopolski et al. 1995); however, in fMRI, k-space coverage with multishot techniques suffer from significantly increased sensitivity to temporal fluctuations induced by physiological processes and motion (Hu and Kim 1994; Hu et al. 1995; Moeller et al. 2006; van der Zwaag et al. 2011). These temporal instabilities can
be minimized by post-processing strategies in multishot as well as in single shot techniques (e.g. Hu and Kim 1994; Hu et al. 1995; Mitra et al. 1995; Biswal et al. 1996; Ogawa et al. 1996; Mitra et al. 1997; Mitra et al. 1997; Mitra and Pesaran 1999; Glover et al. 2000; Pfeuffer et al. 2002; Van De Moortele et al. 2002). However, such strategies require intrinsically good SNR and, in some cases, rapid image acquisition so as to capture these temporal fluctuations accurately.

Of course, high resolution whole brain coverage necessarily increase the volume acquisition times (TR); using segmented, multishot approaches for a single slice increases this time further, leading to prolonged times for collecting sufficient data in an fMRI time series. With the advent of parallel imaging methods, which were discussed previously, the use of partial Fourier sampling (Feinberg et al. 1986), and improved gradient performance, many of the aforementioned technical limitations at high fields have been significantly alleviated. Using these advances, an example of what is feasible at 7 Tesla with GRE EPI is shown in Figure 7; these were images obtained with 0.75 mm isotropic resolution using a reduction factor, R, of 4 (i.e. 4 fold undersampling of the k space, or equivalently, 4 fold reduction in FOV, leading, in this case, to 4 fold maximum aliasing). Despite this, to date only a few whole-brain applications have been presented at ultrahigh fields and have mostly employed large slices, inter-slice gaps or have sacrificed temporal resolution (e.g. Bianciardi et al. 2009; Bianciardi et al. 2009; Bianciardi et al. 2009; Poser and Norris 2009; van der Zwaag et al. 2009; van der Zwaag et al. 2009; Poser et al. 2010), thus not fully exploiting the full benefit of ultrahigh field fMRI.

Although parallel imaging and partial Fourier sampling reduce the number of phase encoding steps for spatial encoding, and, consequently the echo train length after a single RF pulse, they do not necessarily reduce whole brain image acquisition times significantly in many applications. This is because a physiological contrast preparation period must precede the spatial encoding period for each slice and this contrast preparation period can equal or exceed the time employed for the spatial encoding in the EPI echo train. 3D echo volume (EVI) (Mansfield et al. 1994) avoids the repetition of the contrast encoding time by following a single contrast preparation period with subsequent 3D volume coverage in a single echo train. However, this approach has limitations in spatial resolution and image quality due to longer echo trains needed to fully encode
the volumetric spatial information in the relatively short acquisition period dictated by $T_2^*$, especially at high fields; the consequence is distortions and blurring on two of the 3D image axes, as well as a loss in SNR. Multi-shot (segmented with multiple excitation) 3D EPI approaches that have produced high quality images (e.g. Wielopolski et al. 1995; Poser et al. 2010) overcome this limitation; however, in order to achieve very rapid coverage of the whole brain, high parallel imaging reduction factors must be employed in the two orthogonal phase encoding dimensions in the 3D acquisition, leading to significant SNR losses due to the omitted phase encoding steps; these approaches also suffer from increased temporal fluctuations inherent in multishot techniques as discussed above. Echo shifting approaches, PRESTO (Liu et al. 1993; Golay et al. 2000), increase volume coverage efficiency in fMRI by taking advantage of TE delays to apply additional RF pulses, but run into restrictions at higher magnetic fields when $T_2$ and $T_2^*$ become inherently short.

Motivated by the prospect of acquiring high resolution, single shot whole brain images at 7 Tesla, but confronted with the aforelisted limitations in whole brain coverage as a major impediment, we recently introduced multiband (MB) EPI imaging for fMRI based on acceleration in two dimensions, slice acceleration and in-plane phase-encode acceleration simultaneously, in order to achieve significant reductions in whole brain coverage time without substantial or even noticeable degradation in image quality or SNR (Moeller et al. 2008; Moeller et al. 2010). In this technique, several slices are simultaneously excited using multiband RF pulses and subsequently acquired in a single EPI echo train using additional acceleration along the phase encode direction; the simultaneously acquired slices are then unaliased using parallel imaging principles. The acceleration using multiple slice excitation dates back to 2001. Simultaneous acquisition of multiple slices using GRE and subsequently unfolding them using parallel imaging principles was described for spine imaging (Larkman et al. 2001) it 2001. This GRE approach was further pursued in the CAIPIRINHA (Controlled Aliasing In Parallel Imaging Results In Higher Acceleration) (Breuer et al. 2005; Breuer et al. 2006) technique where unaliasing of slices were improved by manipulating the phase of the RF excitation pulses progressively for each k-space line in the GRE sequence, so as to effectively shift the simultaneously acquired slices relative to each other in the phase encoding direction. The fMRI community, including us, had probably ignored these earlier
Two Decades of fMRI

efforts because they were based on gradient recalled echoes with application of one RF pulse per k-space line coverage, an approach that is simply inadequate for fMRI. A method employing gradients blips to achieve a shift for controlled aliasing as in CAIPIRINHA in multiband EPI imaging was also described in an abstract (Nunes et al. 2006). However, the approach led to voxel tilting that was in general not desirable. In our EPI implementation such controlled aliasing was not employed. Yet, we were able to attain 4 fold slice acceleration simultaneously with 4 fold acceleration in the phase encode direction, resulting in 16 fold two dimensional acceleration with 16 fold maximum aliasing.

The multiband/multislice approach has been applied rarely, either in its original form or in the CAIPIRINHA versions; subsequent to our work reviving the approach and demonstrating its potential utility (Moeller et al. 2008; Moeller et al. 2010), its use has been rapidly increasing: It has been incorporated into the Steady State Free Precession (SSFP) sequence (Stab et al. 2011) and radial acquisitions (Yutzey et al. 2011). CAIPIRINHA (Breuer et al. 2005; Breuer et al. 2006) like shifting of slices in Multiband EPI acquisition to improve g-factor penalties in unaliasing, using an gradient blip strategy similar to the strategy described previously (Nunes et al. 2006), but avoiding the voxel tilt, was introduced (Setsomrop et al. 2011). We have described a modification of the Multiband EPI method that combines it with the simultaneous image refocused (SIR) scheme (Feinberg et al. 2002) which temporally interleaves signals from several slices within an EPI echo train; in this new approach, referred to as Multiplexed-EPI (M-EPI), if multiband pulses with m bands are used for each of the s interleaved RF pulses in the SIR technique (where m and s are positive non-zero integers), the result is simultaneous acquisition of m times s (i.e. m • s) slices in a single echo train. The Multiplexed-EPI sequence and some images obtained with the different acceleration factors are shown in Figure 9. We have further optimized acquisition strategies to improve the ability to unalias multiple slices, achieving 6 to 8 fold slice acceleration at 3 Tesla with a 32 channel RF coil (Xu et al. 2012).

Although these slice-accelerated sequences have been introduced for fMRI initially at 7 Tesla (Moeller et al. 2008; Moeller et al. 2010), so far their use to study human brain function has only been reported at 3T (Smith et al. 2012); in this work, taking advantage of the most recent improvements achieved in our lab (Xu et al. 2012), and exploiting the improved temporal resolution and the
large number of data points that can be acquired in a given period of time, time series data obtained with Multiband EPI was employed to identify functionally distinct networks, referred to as *temporal functional modes* (TFM), by virtue of their *temporal independence*; these functionally-distinct modes of spontaneous brain activity were, in general, quite different from resting-state networks (RSN) previously reported, and may have greater biological interpretability. Some of these TFMs were shown to subdivide the default-mode network identified as a RSN.

**FIGURE 9:** The pulse sequence for regular EPI and Multiplexed EPI (M-EPI) and images obtained with Multiplexed EPI images. EPI pulse sequence generates a single slice image during each readout train, which is repeated by the number of slices to scan the whole brain; replacing the single slice pulse with multiband (MB) pulse in this sequence to excite several slices simultaneously, and then unaliasing them using array coil sensitivity profiles yields the Multiband technique (79) which requires fewer repeats to scan the whole brain. Multiplexed-EPI (M-EPI) pulse sequence combines the SIR approach with MB technique: SIR consecutively excites $s$ slices ($s = 3$) is shown in the pulse sequence diagram with pulses in red, blue and green) and acquires them in a single echo train. Using MB pulses that simultaneously excite $m$ slices for each of the single slice pulses in the SIR approach produces the M-EPI sequence, with a *slice acceleration* of $s \cdot m$ leading to $(s \cdot m)$ number of slices collected in a single echo train. Four slices from 2 mm isotropic resolution, 60 slice, whole brain 3 Tesla data set obtained with the M-EPI technique are shown, together with the $(s \cdot m)$ acceleration factors employed. Adapted from Moeller et al. 2010; Feinberg et al. 2010.
Two relatively recent developments with fMRI are worth mentioning briefly with some references given for the reader to pursue further should they decide to do so. These developments involve the use of multivariate machine learning algorithms to decode brain activity patterns and a strategy that can be termed as encoding (Naselaris et al. 2011). Both of these methods offer new possibilities, especially when coupled with the information rich, high resolution, and high specificity functional mapping available at ultra-high magnetic fields.

Multivariate machine learning algorithms applied to GRE fMRI data obtained at 3 Tesla were shown to decode visual stimuli (Haxby et al. 2001; Haynes and Rees 2005; Kamitani and Tong 2005). These algorithms were able to reveal the presence of information that is known to be functionally segregated in cortical columns, e.g. ocular dominance, orientation and direction of motion. This was considered a particularly surprising result since the data were obtained with a large voxel size, 3x3x3 mm$^3$, while these columns have dimensions of ~1 mm; in addition, the data were acquired at 3 T with GRE fMRI where functional specificity is not expected to be sufficient to image at the columnar level. The functional images generated in this process have no resemblance to the actual columnar architecture. Nevertheless, voxels were found that demonstrated a preferential activation corresponding to a unique column, such as an orientation specific column.

The mechanism by which such low-resolution imaging decodes information represented at a much finer scale relative to the voxel size has been a source of debate and experimentation. Biased sampling of cortical columns by the large voxels has been hypothesized (Kamitani and Tong 2005; Haynes and Rees 2006; Kamitani and Tong 2006; Bianciardi et al. 2011). Biased sampling occurs due to the specific position that a large voxel takes within a fine columnar organization with local variations, even if the overall preferences represented by the columns are distributed equally across the investigated cortical region. Alternatively, draining macroscopic blood vessels can have bias with respect to cortical columns and thus could display selective responses (Kamitani and Tong 2005; Kamitani and Tong 2006; Chaimow et al. 2010; Shmuel et al. 2010; Bianciardi et al. 2011). The presence of bias in draining vessels with
respect to ocular dominance columns was directly demonstrated in high-resolution columnar level mapping at 7 Tesla (Yacoub et al. 2007). Further analysis of these data with decoding algorithms revealed that discriminative power was conveyed by both macroscopic blood vessels and coarse-scale organization of the columnar structure, in other words irregularities in the distribution of different columns, in the gray matter areas (Shmuel et al. 2010). A detailed modeling study was performed to examine the latter, considering the pattern of cortical columns, the hemodynamic point spread function at 3 Tesla, the voxel size employed in the 3 Tesla decoding studies, and noise (Chaimow et al. 2011). These model studies demonstrated that sub-voxel supra-Nyquist spatial frequencies, including frequencies near the main frequency of the ocular dominance organization (spatial frequency of 0.5 cycles per mm) cannot contribute to the differential contrast in GRE fMRI and that the differential functional contrast of local origin is dominated by low-amplitude contributions from low frequencies, associated with irregularities of the columnar pattern (Chaimow et al. 2011). However, decoding due to such irregularities predicted decoding performances much lower than those that have been experimentally observed at 3 Tesla, suggesting the presence of additional mechanisms, such as the presence of biased draining veins implicated in Shmuel et al. (2010).

The decoding approaches described above rely predominantly on classification of image voxels where a pattern of activity across multiple voxels is used to construct a discrete class of stimulus response. Recent studies have moved beyond classification and demonstrated stimulus reconstruction to produce a literal picture of the image that was presented. Reconstruction of geometric stimuli composed of flickering checkerboard patterns was achieved by analyzing the responses of voxels in early visual areas using GRE fMRI (Thirion et al. 2006; Miyawaki et al. 2008). The more complex task of reconstruction of natural images was undertaken with significant success using forward encoding models of each voxel (Kay et al. 2008; Naselaris et al. 2009; Naselaris et al. 2011). These results suggest that it may soon be possible to reconstruct a picture of a person’s visual experience from measurements of functional imaging. In this quest, ultrahigh field fMRI is bound to play a critical role since it will provide many more informative and independent voxels with finer, more specific, information content so as to enable the expansion of encoding models and provide substantial increases in discriminative power.
CONCLUSIONS

Our understanding of MR detectable functional signals in the brain has improved significantly over the years since the development of the methodology. At the same time, we have seen major advances and refinements in instrumentation, such as the introduction of ultrahigh field instruments of ever increasing sophistication and refinement, and data acquisition and image analysis methodologies that have significantly impacted data quality. These fantastic improvements today enable mapping of cortical columns, layer specific activation or brain reading. These advances are sure to continue unabated for years to come as new instruments like the 10.5 Tesla system expected at the University of Minnesota and the 11.7 Tesla systems planned for the Intramural Research division of NIH and NeuroSpin Laboratory in France will come on line. With these instruments and the ever improving data analysis methods, MR based functional imaging may be performed in the not so distant future with totally different approaches and provide substantially better information than the already available rich techniques at the cutting edge today.

APPENDIX 1: EXTRAVASCULAR BOLD EFFECT

If one considers an infinite cylinder as an approximation for a blood vessel with magnetic susceptibility difference $\Delta \chi$, then the magnetic field at any point in space, will be perturbed from the applied magnetic field (Springer and Xu 1991). Inside the cylinder, the perturbation, $\Delta \nu$, in Hz will be given by the equation

$$\Delta \nu^{in} = \Delta \chi \epsilon_0 (Y_0 - Y)(\gamma B_0)(\cos^2 \theta - 1/3)$$  \hspace{1cm} (1)

At any point outside the cylinder, designated by the vector $\vec{r}$, the magnetic field will vary depending on the distance and orientation relative to the blood vessel and the external magnetic field direction, according to the equation:

$$\Delta \nu^{out}(\vec{r}) = \Delta \chi \epsilon_0 (Y_0 - Y)(\gamma B_0)(r_b/|\vec{r}|)^2 \sin^2 \theta \cos(2\phi)$$  \hspace{1cm} (2)

In these equations, the parameter $\gamma B_0$ is the Larmour frequency of water protons expressed as an angular frequency in rad/s, $B_0$ is the
magnetic field and $\gamma$ is the gyromagnetic ratio, which is $2.6751965 \times 10^8 \text{ rad s}^{-1} \text{ Tesla}^{-1}$ or equivalently $42.577$ in units of MHz/Tesla in a spherical water sample. The angles are defined in Figure 10; $\theta$ is the angle the cylinder makes the magnetic field $B_0$, generated by the magnet; for any given point in space, $\phi$ is the angle of the vector from that point running perpendicularly to the cylinder and the plane defined by the cylinder and the magnetic field $B_0$. $\Delta \chi_0$ is the maximum susceptibility difference, expressed as parts per million (ppm) expected in the presence of fully deoxygenated blood (and depends on the hematocrit); $Y_0$ is the fraction of oxygenated blood at which the susceptibility difference between the blood and the surrounding tissue is zero, and $Y$ is the fraction of oxygenated blood present in the vessel under consideration. $Y_0$ is usually assumed to be 1, as in (Ogawa et al. 1993); recent data assigns it a value of ~0.95 (Spees et al. 2001). Also varying values exist for $\Delta \chi_0$; 0.264 ppm at unit hematocrit was recently determined experimentally (Spees et al. 2001) and differs from the value of 0.18 and 0.31 ppm used in some of the other earlier studies. At a hematocrit value of 0.4, $\Delta \chi_0$ is, therefore, (0.264•0.4), which is 0.106; a value of 0.1 was used in (Ogawa et al. 1993). The parameter $r_b$ designates the cylinder radius, $r$ is the distance from the point of interest to the center of the cylinder in the plane normal to the cylinder. Note that outside the cylinder, the magnetic field changes rapidly over a distance comparable to two or three times the cylinder radius; at a distance equal to the diameter of the cylinder from the cylinder center, $\Delta \chi_{\text{out}}$ is already down to 25% of its value at the cylinder boundary.

Since there are often different units and nomenclatures used, we can present a sample calculation: Using $\Delta \chi_0$ of 0.264 ppm at unit hematocrit, a blood hematocrit of 0.4, $Y_0$ of 0.95, the parameter $\Delta \chi_0(Y_0 - Y)(\gamma B_0)$ is calculated for 1 Tesla and at $Y$ of 0.6, typical of brain venous blood, as $(0.264 \times 10^{-6} \times 0.4) \times (0.95 - 0.6)(2.6751965 \times 10^8)$ = 9.89 Hz; thus, $\Delta \chi_0(Y_0 - Y)(\gamma B_0)$ ~10 Hz for 1 Tesla, an easy round number to remember. Note that this applies for the veins. Capillaries, which also contain deoxyhemoglobin, do not have $Y = 0.6$. In these vessels $Y$ changes from arteriolar level (~1) to the venous level (~0.6) along the length of the blood vessel. Thus, using the average along the length of the capillary, $\Delta \chi_0(Y_0 - Y)(\gamma B_0)$ is ~5 Hz for 1 Tesla, half of the value for veins.
APPENDIX 2: BRIEF INTRODUCTION TO DATA ACQUISITION IN FMRI:

In BOLD contrast, a signal is acquired simply after a delay, TE, following excitation. A refocusing pulse may or may not be applied during this delay. When refocusing pulses are not applied, one uses field gradient reversal to form an echo in imaging; consequently, such echoes are called gradient-recalled or gradient echoes, abbreviated simply as GRE. This is the first MR sequence used for functional mapping with the BOLD approach (Bandettini et al. 1992; Kwong et al. 1992; Ogawa et al. 1992). In the presence of a refocusing pulse, the spin-echo (SE) signal (i.e. echo-amplitude) detected after the echo time TE decreases according to $\exp(-\text{TE}/T_2)$ where $T_2$ is the spin-lattice relaxation time associated with the loss of magnetization with time in the transverse plane, the plane perpendicular to the main static magnetic field $B_0$. Spin echoes can also yield functional information in the brain leading to SE based fMRI. SE fMRI is sensitive only to a subset of the processes that lead to BOLD contrast in the GRE experiment, while everything that provides functional signals in SE fMRI also contribute to GRE fMRI.

In a gradient echo, signal loss occurs through $T_2$ relaxation as well as by signal cancellation that arises due to dephasing of the magnetization in the presence of magnetic field inhomogeneities.
The latter is recoverable with a refocusing pulse and does not contribute to spin-echoes. In a gradient echo image, the appropriate relaxation constant of magnetization in the transverse plane is $T_2^\ast$. $T_2$ of pure water is dominated by the magnetic interactions between the two hydrogen atoms, through the dipole-dipole coupling. In the human brain, this is still the case but the water is not in a pure form; rather, it exists in many different environments, such as tightly or weakly bound to macromolecules and metal ions, or free of such binding. The presence of these different states and the exchange between them affect the $T_2$ of the water resonance. The hydrogen atom of the water molecule itself can exchange between water and exchangeable $-NH$ or $-OH$ hydrogens on macromolecules. All of these processes affect the water proton $T_2$ and impart the contrast that allows the differentiation of one type of tissue from another in MRI. Another mechanism that contributes to $T_2$ is diffusion of water molecules around magnetic field inhomogeneities; normally in the brain this is not a very large effect though it is present and detectable especially at high magnetic fields (Bartha et al. 2002; Michaeli et al. 2002).

$T_2^\ast$, on the other hand, has contributions from all of the process that affect $T_2$ plus the dephasing effect in the presence of magnetic field inhomogeneities. In a container of pure water placed in a fully homogenous external magnetic field, $T_2$ equals $T_2^\ast$, except near the boundaries of the container where the magnetic field can be non-uniform. This is not the case in the brain because of magnetic field inhomogeneities that span scales ranging from microns to many millimeters and even centimeters as encountered around the air filled cavities adjacent to the brain. On the micron to millimeter scale, the inhomogeneities exist around metal containing enzymes, intracellular structures, myelin, and deoxyhemoglobin containing blood vessels ranging from capillaries to large veins; it is this last effect that provides functional mapping signals. Oxygenated blood is diamagnetic and has similar properties to tissue. However, the deoxyhemoglobin molecule is paramagnetic, and its presence leads to a large susceptibility difference between compartments that contain this molecule and others that are devoid of it. Thus, against a backdrop of numerous contributions to the inherent $T_2$ and $T_2^\ast$ of brain water, BOLD contrast refers to a $T_2^\ast$ or $T_2$ mechanism arising from magnetic field inhomogeneities generated by magnetic susceptibility differences across the luminal boundaries of blood vessels. During increased neuronal activity, deoxyhemoglobin con-
tent is altered in the brain; this change is accompanied by small but measurable decreases in $T_2$ and $T_2^*$, which was experimentally documented in the early days of fMRI (Menon et al. 1993).

Image acquisition can be accomplished using slices where a frequency selective RF pulse is employed to restrict the signal origin to a single slice, typically ~1 to 3 mm in thickness at the present time, though as large as 10 mm slices were not uncommon in the early days of fMRI.

In MR imaging terminology, k-space refers to the two or three-dimensional matrix of data points collected during image acquisition. When the image acquired is from a slice only, the k-space data reside in two dimensions, representing the encoding along two orthogonal directions that define a plane perpendicular to the slice selection direction. A 2-dimensional Fourier transform then converts this into an image of the slice. In techniques like FLASH (Haase et al. 1986, also see reviews Haacke and Tkach 1990; Chien and Edelman 1991; Haacke et al. 1991), one line along a single dimension of k-space is collected after each RF pulse and multiple RF pulses are employed to cover the entire k-space in seconds. In approaches like single-shot Echo Planar imaging (EPI) (Mansfield 1977) or SPIRAL imaging (e.g. Yang et al. 1998), the entire k-space is covered after a single RF pulse using a train of gradient echoes in ~20 to 100 ms depending on the available hardware and the field magnitude at which the data are collected. Segmented EPI schemes (McKinnon 1993; Feinberg and Oshio 1994; Wielopolski et al. 1995) cover more than 1 line but less than the entire k-space in a single echo train after one RF pulse. They have been particularly useful in high-resolution functional imaging studies (Yacoub et al. 2007; Yacoub et al. 2008) because the higher resolution dictates the acquisition of many more k-space lines and many more points along each k-space line.

Single shot techniques provide distinct advantages to the others for fMRI. These include the ability to cover the whole brain rapidly, and to partially suppress temporal signal fluctuations in an fMRI time series. The latter touches on a very interesting subject: Namely, in consecutively acquired images from the brain, temporal signal variations contain contributions from and can even be dominated by physiologically induced processes such as blood vessel pulsation (vasomotion), respiration, and spontaneous neuronal activity. Of course, the intrinsic signal-to-noise ratio (SNR) of a single image has to be sufficiently good to detect these physiologically in-
duced variations. These temporal instabilities come into single shot versus multiple shot acquisitions in a different way and are much more deleterious for the latter (Hu and Kim 1994; Hu et al. 1995). They can be minimized by post-processing strategies in multiple shot as well as in single shot techniques (e.g. (Hu and Kim 1994; Hu et al. 1995; Mitra et al. 1995; Biswal et al. 1996; Ogawa et al. 1996; Mitra et al. 1997; Mitra and Pesaran 1999; Glover et al. 2000; Pfeuffer et al. 2002; Van De Moortele et al. 2002)). However, such strategies require intrinsically good SNR and, and in some cases, rapid image acquisition so as to capture these temporal fluctuations accurately.

At higher magnetic fields, it is necessary to acquire single shot images faster because the MR signal disappears faster subsequent to excitation (i.e. the $T_2^*$ is shorter) due to increased magnetic field inhomogeneities. A critical development in this respect has been parallel imaging (PI) techniques which utilize the spatial information inherent in multiple surface coils distributed over the sample so as to omit some of the k-space lines as if the image was being acquired with a reduced field of view (FOV). Consequently, the time spent acquiring the signal after the RF pulse is shorter, which translates into shorter echo trains for EPI. When processed without the use of the spatial information from array coil sensitivity profiles, images acquired in this fashion display sections that are folded on top of each other, in other words sections that are aliased. However, they can be unfolded using parallel imaging techniques. The utility of parallel imaging has been examined for functional imaging early on after their introduction (de Zwart et al. 2006; Moeller et al. 2006); today, they are typically part of the normal fMRI acquisition, especially at high magnetic fields.

Parallel imaging works better at high magnetic fields (Ohliger et al. 2003; Wiesinger et al. 2004; Wiesinger et al. 2006) and high magnetic fields need the use of parallel imaging relatively more because of the short window available for data acquisition due to the short $T_2^*$ (e.g. see (Yacoub et al. 2001; Yacoub et al. 2003; Uludag et al. 2009) and references therein). Thus, parallel imaging and high fields are synergistic. Partial Fourier (Feinberg et al. 1986) sampling, or sparse data sampling approaches (Liang et al. 2003) are also approaches that reduce the image encoding times. With the use of these techniques, it is feasible today to acquire single shot high-resolution EPI images at high fields such as 7 Tesla, as illustrated in Figure 7.
Parallel Imaging, and Partial Fourier sampling reduce the number of phase encoding steps for spatial encoding, and consequently, the echo train length after a single RF pulse; however, they do not necessarily reduce whole brain image acquisition times significantly in many applications. Much faster acquisitions can be obtained using the Multiband technique as discussed in the main body of the article.

ENDNOTES

1. The frequency range 300 MHz to 3 GHz is defined as Ultra High Frequency (UHF) (see http://en.wikipedia.org/wiki/Ultra_high_frequency). The hydrogen nucleus resonance frequency at 7 Tesla is ~300 MHz i.e. in UHF band. Therefore, 7 to 70 Tesla is defined as Ultra High Field (UHF).

2. In reality, this statement is correct strictly if we treat blood as a homogeneous medium with a magnetic susceptibility that differs from the surrounding tissue. However, in blood, deoxyhemoglobin is compartmentalized within red blood cells. This leads to magnetic field inhomogeneities around the red blood cells, thus to a non-uniform field distribution within the blood. However, the spatial scale of these inhomogeneities is sufficiently small that water diffusion averages them out; thus, blood water experiences effectively a time-averaged uniform field.

3. The resonance frequency, also known as the Larmor frequency, of a magnetic nucleus such as that of hydrogen atom, is given by $\gamma B_0$ where $\gamma$ is the gyromagnetic ratio which is a fundamental constant associated with the nucleus and $B_0$ is the magnetic field experienced by the nucleus.

ACKNOWLEDGEMENT

The work reported in this review that was carried out at the Center for Magnetic Resonance Research (CMRR), University of Minnesota, was supported by a Biomedical Technology Research Center grant (Grant Number P41 RR008079 from NCRR/P41 EB015894 from NIBIB) and Neuroscience Center Core grant P30 NS057091 from NINDS of the National Institutes of Health, and the Keck Foundation.
REFERENCES


