Laminar fMRI at ultra-high fields

Over the past decade, 7 T has grown from being a research-only platform that provides structural and functional images of the human body with an unparalleled amount of detail and contrast, to a clinical platform that provides improvements made to patient diagnosis and intervention with reduced scan durations. Recently, the improvements have allowed Siemens to receive CE and FDA certifications for their 7 T Terra system for clinical use. Functional MRI researchers, typically neuroscientists with a cursory education in MR-physics, are increasingly switching to 7 T given its significant advantages over conventional field strengths ($\leq$ 3 T). Currently, the range of neuroscientific questions that can be addressed is becoming larger ($\approx$ laminar/columnar circuitry), and the scale of cortical organisation and function becoming smaller ($\approx$ sub-millimetre, sub-second regimes). While it is still possible for plug-and-play fMRI with standard coils and resolutions at 7 T, similar to that at 3 T, high-resolution fMRI comes with a lot of challenges that can be addressed to a large extent, by making informed choices regarding hardware, sequences, acquisition types and analysis strategies. Over the course of the studies reported in this thesis, we investigated different high-resolution anatomical and functional acquisition strategies, alternative contrast mechanisms and analysis approaches for laminar fMRI applications at ultra-high field ($\geq$ 7 T). In this chapter, we discuss the high-resolution fMRI acquisition, processing and analysis strategies that were adapted and developed over the course of this thesis and conclude with a summary.

5.1 Data acquisition and processing

A researcher having a well-controlled experimental design investigating laminar circuitry at 7 T using fMRI is faced with an overabundance of options when it comes to acquiring data. Decisions taken for the data acquisition are driven pri-

marily by previous work reported in literature (most often at 3 T or lower) and legacy data analysis approaches which should "just work". The first step in this process and an often under-appreciated aspect of an fMRI study is choosing the right hardware. Almost every high-field site has specialised multi-receive array coils focussed on specific brain regions such as visual, auditory cortices etc. For example, if the research question pertains to the early visual cortex, the default is to choose the whole-brain coil (despite the less densely packed coil loops near the occipital lobe), even though the high-resolution functional data being acquired is of partial brain coverage i.e. limited to the early visual cortex. This choice is driven by convenience of conventional data processing tools which were for whole-brain analysis by design. However, it would be significantly advantageous, at this juncture, to consider the availability of specialised region-of-interest coils (such as, [197] used in **Chapter 3**) as they have densely packed coil arrays around the region-of-interest that enable achieving higher spatial resolution, acceleration factors and (t)SNR than the default whole-brain coils [263]. The visual coil, for instance, also has the additional advantage of having a greater field-of-view for stimulus presentation than its whole-brain counterpart. This comes at the cost of using new and unconventional data processing strategies which are often not well documented and have several additional compiler and library dependencies. Nevertheless, the advantages do outweigh the disadvantages of using specialised hardware and processing the data using new, command-line driven, unconventional tools are quite manageable. The next step is choosing the appropriate MR-sequences for data acquisition.

5.1.1 Anatomical

Choice of anatomical imaging sequence typically providing a T_1 and/or T_1-weighting for improved GM-WM delineations (e.g. MPRAGE [116], MP2RAGE [117], MEMPRAGE [264], MI-EPI [131]) per se may not very critical but depending on the sequence, it can require additional post-processing in order to be useful. Please note, several laminar fMRI studies acquire anatomical data at 3 T (see 1.1), this is but an example of the aforementioned legacy acquisition schemes (circa 2010). This is no longer a requirement as improved bias-correction algorithms such as N4 [265] for traditional MPRAGE (circumventing the need to acquire additional proton-density weighted images) or the (nearly) bias-free MP2RAGE are preferable. Since the challenges of signal inhomogeneities at ultra-high field can now be dealt with in post-processing, acquiring anatomical data at a lower field strength (in many cases, a luxury) need not be considered. An additional reason to avoid across field strength data acquisition, if avoidable, is that it comes with much more challenges such as requiring gradient non-linearity corrections and across-session co-registrations.

Despite its advantages, one of the reasons why MP2RAGE utilisation is often met with hesitance is that pre-processing algorithms developed for MPRAGE-like acquisitions do not work out of the box. For instance, the MP2RAGE T_1-w as it comes out of the scanner has a very noisy non-brain background (Fig. 5.1, left) and skull-stripping with conventional tools such as *bet* fail. ANTs and SPM offer strategies which are slightly convoluted workarounds. However, the author of the MP2RAGE paper has a simple script available on his Github page[1], that gets rid of this by including a regularisation term to the image reconstruction and renders it very similar to an MPRAGE image(Fig. 5.1, right), and all the conventional tools

[1] `https://github.com/JosePMarques/MP2RAGE-related-scripts`

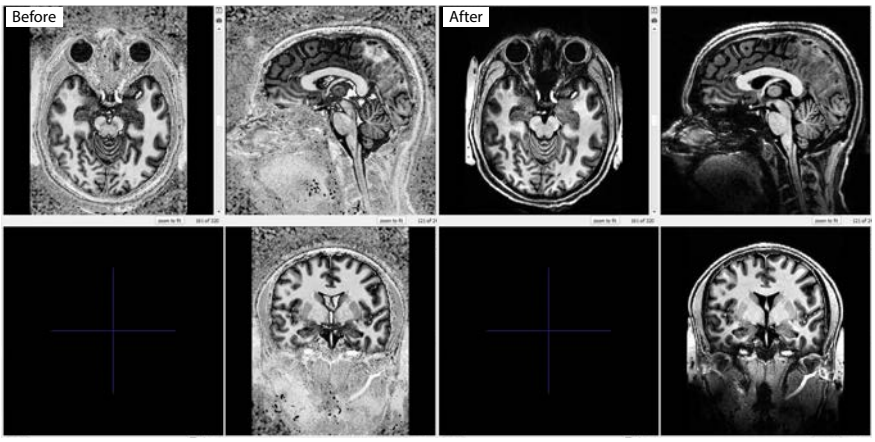

Figure 5.1: The background noise issue with MP2RAGE T_1-w image as obtained by default (left) and the regularised reconstruction using code by José Marques. (`https://github.com/JosePMarques/MP2RAGE-related-scripts/blob/master/DemoRemoveBackgroundNoise.m`)

just work. Alternatively, a simpler processing step of multiplying the MP2RAGE T_1-w by the second inversion image ($\approx$ proton density like contrast) also yields a similar result albeit requiring further intensity inhomogeneity correction. Another advantageous aspect of the MP2RAGE in this regard would be that the first inversion image is specifically optimised to have increased GM-WM contrast and the quantitative T_1 map has increased CSF-GM contrast. Since all these images are in the same space, the two boundaries for laminar analysis can be extracted from these separate contrasts.

In the studies reported in this thesis, we decided to opt for the MP2RAGE sequence for a couple of reasons. First, the additional quantitative T_1 map computed in the MP2RAGE was important for the T_1 comparisons made in **Chapter 2** and errors in T_1 estimates due to $B_1{}^+$ could be corrected post-hoc using an Sa2RAGE sequence. We show that these corrected-T_1 estimates are in the same range as

those estimated from a multiple inversion-recovery EPI (MI-EPI) sequence. Second, due to our use of a specialised visual cortex coil in **Chapter 3** the Sa2RAGE-correction proved critical to correct the strong B_1 in the occipital pole to aid better segmentation. In **Chapter 4**, having the additional contrasts such as the quantitative T_1 image also allowed us to adapt the Freesurfer algorithm for pial surface optimisation designed to use T_2-w images (e.g. SPACE). Although we have not tried using a multi-echo MPRAGE (MEMPRAGE [264]), it would be very interesting to see how the different contrasts can be useful, especially in conjunction with probabilistic frameworks (such as SPM's Unified Segmentation [145]) which can use the information content from multi-contrast data to aid better image segmentation.

5.1.2 Functional

The choices for high-resolution functional acquisition using $T_2{}^*$-weighting are FLASH [198], 2D-EPI [14] and 3D-EPI [141]. While FLASH has the advantage of having little-to-no geometric distortion due to B_0 non-uniformity, it has poor temporal resolution [67] (e.g. one slice $\approx$ 1.76 s in **Chapter 3**). Therefore, for **Chapters 2** and **4**, it was clear we would use an EPI sequence, the decision was between 2D- or 3D-EPI for a nominal voxel resolution of 0.7 mm isotropic. After having carried out pilot experiments, we concluded that in thermal-noise dominated sub-millimetre voxel regimes, the 3D-EPI provided better tSNR and directly translated into more robust functional activation maps. This was later confirmed more systematically by Huber and colleagues [83] (Fig. 5.2) for both the BOLD and CBV-weighted signals using the SS-SI-VASO sequence. Therefore, for the EPI-based datasets in this thesis, we opted for the 3D-EPI acquisition. Additionally, Lutti and colleagues [266] have demonstrated that even for physiological-

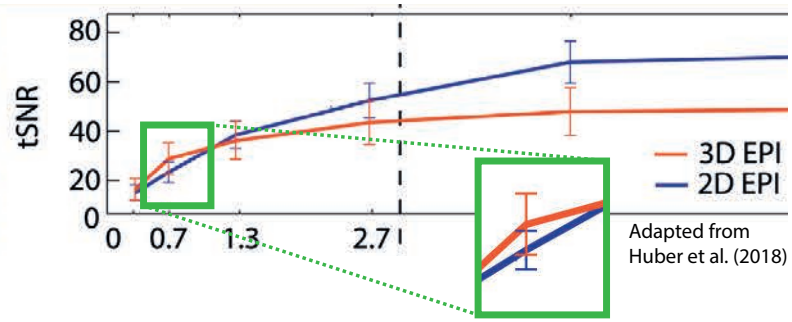

Figure 5.2: Plot of tSNR change with respect to voxel size using 2D- and 3D-EPI readouts (adapted from [83]). The zoomed in box highlights the sub-millimetre regime typically used for laminar fMRI.

noise dominated supra-millimetre regimes, correcting for physiological noise results in $\geq$ 25% increase in tSNR and better sensitivity than 2D-EPI. Our experience does not make the 3D-EPI prescriptive in any sense for all fMRI studies and we just note the importance of pilot experiments before choosing the optimal functional sequence for a study. Dealing with partial-coverage functional data in high-resolution fMRI is especially difficult with conventional data analysis strategies. Existing tools do not manage to deal with these datasets despite there being options to optimise these algorithms for such datasets. For example, something as simple as brain extraction (skull stripping) in FSL's *bet* has a special flag "-Z" when there are few slices in the z-direction. However, despite optimising parameters, it produces erroneous results (Fig. 5.3, left). Brain masking is quite crucial in high-resolution studies because we want to limit the motion-correction/co-registration cost-function minimisations to the brain (even better if using a region-of-interest mask [82]). Here, ITK-SNAP [147] can be particularly useful. It's a highly versatile tool with an excellent set of semi-automated visually-guided segmentation algorithms, making it extremely intuitive to use. The clustering function in partic-

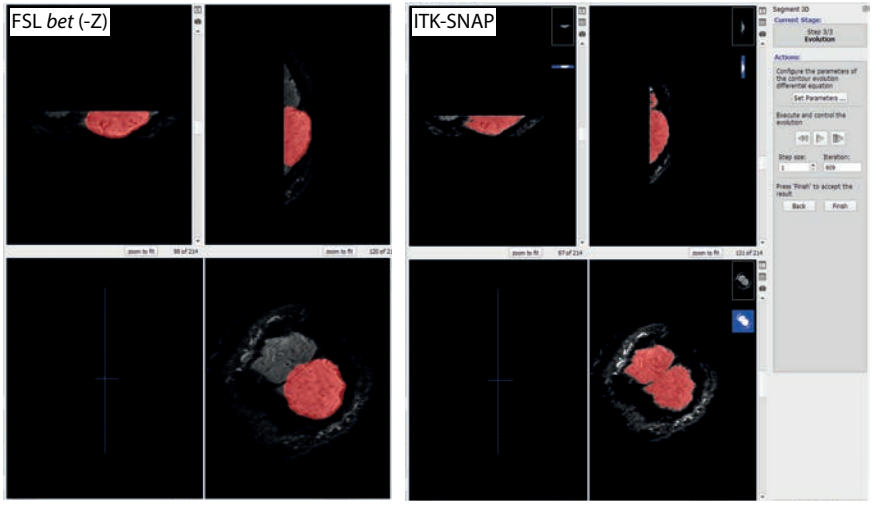

Figure 5.3: Examples of automatic skull-stripping/brain-extraction using FSL *bet* of partial coverage data and the same using ITK-SNAP.

ular can generate a highly accurate brain mask for partial-coverage data (Fig. 5.3, right) in $\leq$ 30 s. ITK-SNAP was essential to create the initial alignment matrices for accurate co-registrations using the ANTs framework [184, 207] as in **Chapters 4** and **5**.

The fact that the anatomical and functional datasets are differently-distorted presents unique challenges. We address this topic exhaustively in **Chapter 2**. Furthermore, we show that it is feasible to use the distortion-matched MI-EPI quantitative T_1 mapping approach with a specialised visual cortex coil in **Chapter 3**. This approach has picked up a lot of traction over the years since its publication and distortion-matched acquisitions have been made available with a 3D-EPI implementation called T123DEPI [267] which utilises an MP2RAGE-like reconstruction scheme.

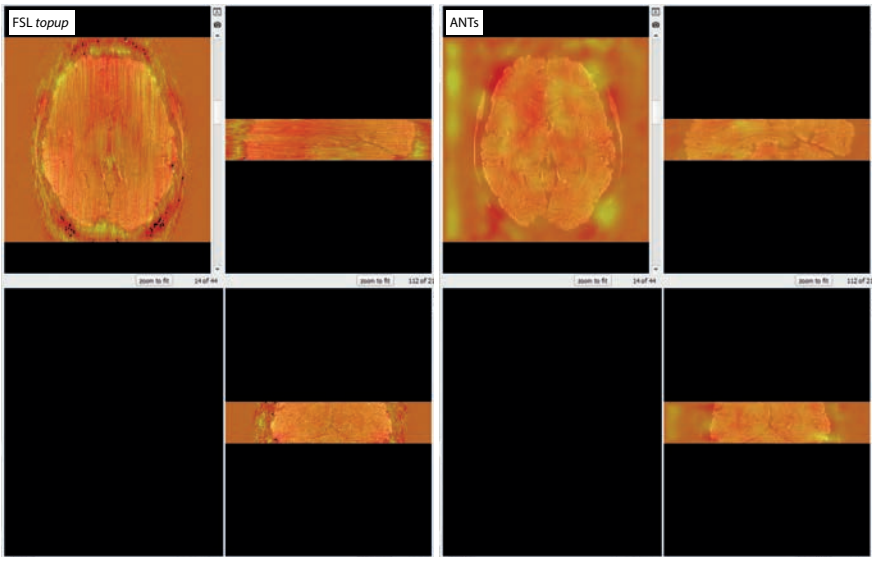

Figure 5.4: Voxel displacement maps from obtained from FSL *topup* and that from ANTs using a constrained SyN deformation.

In the cases where acquiring distortion-matched anatomical datasets was not possible (for e.g. due to time limitations) as in **Chapter 4**, we show that we can still minimally process the functional data (as espoused in **Chapter 2**) by adapting the tools available in the ANTs framework to apply motion-correction, distortion-correction and co-registration to anatomical image in a single resampling step. The voxel-displacement maps by FSL *topup* (Fig. 5.4, left) are very jagged and appear to be slightly lower resolution than those estimated using ANTs (Fig. 5.4, right). This is despite the fact that the FSL *topup* configuration file (specifically - *warpres*) was modified to account for the high-resolution data. Furthermore, ANTs can assess non-linear deformations along all axes, therefore, care was taken so as to ensure that the deformation step were restricted to the phase-encode axis only. The higher resolution warps estimated by ANTs, therefore, result in relatively less

blurring in the images than with *topup*, therefore, likely preserving some of the acquired data's fidelity.

5.2 Laminar analysis

After having performed optimal processing of the anatomical and functional data, the next step is sampling the fMRI signal/statistical values across the cortical depths. An arbitrary number of cortical depths can be defined between the CSF-GM and GM-WM bounds using either an distance-preserving (equi-distant) or volume-preserving (equi-volume) models (Fig. 5.5a, upper-panel). Following this, the fMRI signal can be resampled at these depths in order to achieve a finer separation of the laminar signals than the relatively coarser resolution of the data. Waehnert and colleagues [70] showed that the distance-preserving layering model is sub-optimal in regions of with high-curvature, whereas a volume-preserving layering model better follows the cyto-architecture. However, in our experience (data not shown, **Chapter 2**) at voxel resolutions $\approx$ 0.7-0.8 mm isotropic, and averaging over a region-of-interest did not reveal a significant difference between the two models for laminar sampling [75, 76]. Additionally, the equi-volume approach was infrequently used as only CBS-Tools [202] offered the algorithm bundled within a software package called MIPAV. Recently, due to its open-source nature, there are now community-driven scripts to use the equi-volume model with conventional tools such as Freesurfer [252] and the whole CBS-Tools algorithms have been repackaged in python as *nighres*. Once the layering is carried out, the depth-dependent analysis is then done by resampling the fMRI signal onto dense meshes [66], binning voxels in an upsampled resolution [98] or using regular Cartesian grids [99]. Data resampling in this manner can result in spatial blurring, signal leakage across layers and reduced specificity [72].

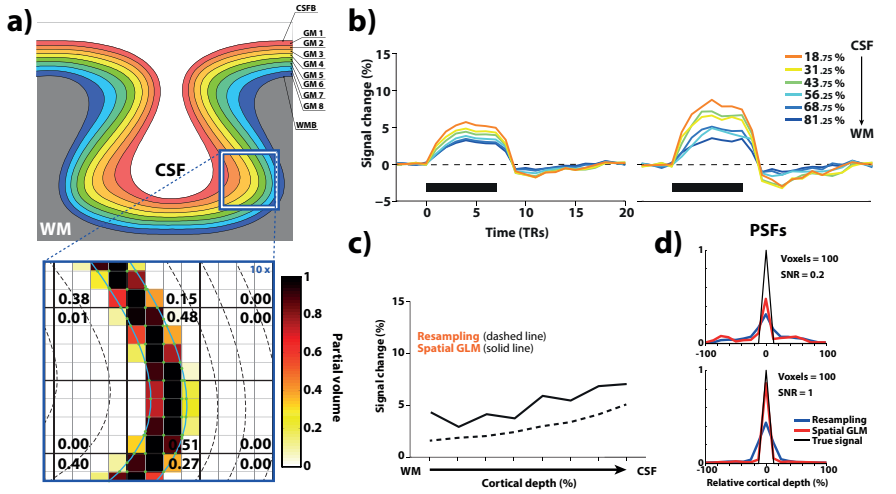

Figure 5.5: (a) top: equivolume layering on a curved model of cortex, bottom: a snapshot of the estimated the laminar-partial-volume-matrix. (b) Laminar time-courses by resampling (left) and spatial GLM (right). (c) Laminar profiles from the two sampling methods. (d) estimated PSFs for the sampling methods under realistic (SNR=0.2) and ideal (SNR=1) scenarios.

An alternative has been proposed to unmix the laminar signal using a spatial general linear model (spatial GLM) [72, 74, 182]. In this approach, a laminar-partial-volume matrix is estimated (Fig. 5.5a, bottom-panel) i.e., the space between two laminar surfaces defines the partial volume contribution of each layer to a voxel that spans several layers. Using data from **Chapter 2**, Havlicek and colleagues [72] have demonstrated the differences between the two approaches (Fig. 5.5b) on the sampled laminar time-courses and the corresponding laminar profiles (Fig. 5.5c) with the resampling approach providing smoother profiles whereas the spatial GLM has more structure to it while exhibiting a similar general trend. Figure 5.5d shows the estimated point-spread functions from simulated data by Havlicek and colleagues [72] for a realistic scenario (SNR=0.2) and an idealistic scenario (SNR=1), showing that in both cases, the spatial GLM approximates more

closely to the ground-truth than the resampling approach. Nevertheless, this approach still requires systematic testing and validation.

Lastly, almost every laminar fMRI paper issues a disclaimer regarding the use of "lamina" to define computationally generated depths and not directly related to cytoarchitectonic layers. While there are plenty of arguments why the word "lamina" is not the sole custody of histologists (discussed at length in *layerfmri.com*), the term's usage in fMRI is still not mainstream enough to merit its use without this disclaimer. A quick, informally conducted twitter poll (Fig. 5.6), revealed that *laminar fMRI* is preferred by a small majority followed closely by *depth-dependent fMRI* and *layer fMRI* despite there being no clear consensus. Nevertheless, this is the hallmark of an evolving field of research and such a high-degree of community engagement is definitely a reason for optimism.

Figure 5.6: Results of a poll conducted on twitter (`https://twitter.com/srirangakashyap/status/1102146422592491521`)

5.3 Vascular bias in GE-BOLD

As introduced briefly in **Chapter 1** and also addressed to some extent in **Chapter 2**, the GE-BOLD contrast used for the vast majority of laminar fMRI studies suffers from spatial confounds associated with intra-cortical ascending veins and draining pial veins. While the some of the pial veins can be spotted with high-resolution anatomical scans, the ascending veins are very difficult to resolve at the voxel resolutions possible thus far. These ascending veins collect

the deoxy-Hb from the micro-vasculature at the site of neuronal activation and transports it to the pial vein at the surface. The distribution of these ascending veins and the accumulative effect of the deoxy-Hb introduces local biases in amplitude and localisation of the laminar BOLD signal in the cortex. There have been modelling efforts to address this intra-cortical drainage effect of the ascending vein [109, 110] and till date, only Marquardt and colleagues [195] have applied the steady-state Markuerkiaga model [110] to high-resolution fMRI data to deconvolve the ascending vein effect from their sampled laminar BOLD profiles. On the other hand, the Heinzle model [109] is a two-depth extension of the balloon model [268] within the dynamic causal modelling (DCM) framework [269] and has been recently improved upon by Havlicek and Uludağ [270] into a multi-compartment laminar BOLD model that is derived using mass conservation principles. Because the Havlicek-Uludağ laminar BOLD model is a fully dynamic model described by differential equations, it allows simulating steady-states and entire time-courses together with all its transients, such as response peak, early-overshoot, post-stimulus undershoot (PSU) or "the elusive" initial dip [271].

5.4 Non-BOLD alternatives

As discussed above, there have been significant strides made in addressing the limitations of the BOLD contrast using model-driven computational approaches and their application and validation will be critical as BOLD remains the contrast-of-choice for many studies given its unrivalled sensitivity. However, there have also been significant developments made in making the non-BOLD functional contrasts such as the SS-SI-VASO (measuring CBV and BOLD) [180], 3D-EPI pulsed ASL (measuring CBF and BOLD) [238] (**Chapter 4**) feasible at ultra-high field

and at sub-millimetre resolutions. This is particularly useful for laminar fMRI as both VASO and ASL offer better spatial specificity than BOLD but have limited sensitivity [80], with most applications arising from sensory cortices with stimuli that elicit very robust activation. Of the available alternatives, VASO has recently seen application to the dorso-lateral pre-frontal cortex (dlPFC) [272] apart from several laminar fMRI studies already carried out in the motor cortex [82, 166, 180]. On the other hand, CBF mapping (ASL) by virtue of the signal is inherently limited in SNR, even compared to total CBV mapping (VASO). This has seen limited application outside of clinical CBF mapping to high-resolution studies let alone fMRI. However, following significant sequence development and optimisation over the course of this thesis by Ivanov and colleagues [84, 88, 238, 273], we demonstrate, for the first time, the feasibility of using sub-millimetre spatial resolution 3D-EPI pulsed ASL for laminar fMRI in the visual cortex (**Chapter 4**). We hope to explore the full potential of the CBF contrast for laminar fMRI in future follow-up studies.

5.5 Concluding remarks and outlook

Since 2010, the field of laminar fMRI has grown at an extraordinary pace together with the increasing adoption of UHF MRI scanners. It is no surprise that one of the principal MRI journals, *Neuroimage*, came out with three Special Issues[2] dedicated to developments encompassing optimised MR-sequences, post-processing algorithms, analysis tools and applications to cognitive neuroscience. **Chapter 2** of this thesis emphasised minimal processing of functional data, demonstrated the feasibility of distortion-matched anatomical acquisitions to aid laminar analysis

[2] 1. *Pushing the spatio-temporal limits of MRI and fMRI* (Eds. Essa Yacoub, Larry Wald), 2. *Neuroimaging with Ultra-high Field MRI: Present and Future* (Eds. Jonathan Polimeni, Kâmil Uludağ), 3. *in progress* (Eds. Jonathan Polimeni and David Norris)

in EPI space, raised the issue of co-registration inaccuracies even with distortion-correction and showed the impact of conventional data processing strategies on the observed laminar profiles. The first laminar fMRI studies using EPI were carried out at 1 mm isotropic resolution [66], many following studies pushed this to 0.8 mm [99] and to 0.7 mm [89, 98]. While we do not know what the limits for spatial resolution are, what is clear is that there is no substitute for higher resolution. One way to push the boundary was inspired by line-scanning [79] animal fMRI by favouring ultra-high resolution along the dimension-of-interest (layers or columns) as we demonstrate using anisotropic voxels in **Chapter 3**. Future applications of this technique should incorporate some form of prospective motion correction either using optical-tracking, field-monitoring or even passively using custom designed head-cases[3]. Some preliminary results demonstrating the feasibility of a multi-echo version of the anisotropic FLASH fMRI in the motor cortex at 9.4 T was presented recently [274]. At ISMRM 2019, research groups in Boston [275] and Utrecht [276] have taken our approach further by actually implementing animal line-scanning like schemes for diffusion and fMRI respectively. This *line* of research is slowly, but surely gaining traction with many of the technological and analysis kinks to be addressed in the upcoming years. Finally, a take-away from a review of the studies in the field so far is that the functional neuroimaging community is now spoilt for choice when it comes to carrying out a laminar fMRI study. An example is that only a few years ago, the contrast of choice for high-resolution laminar fMRI in humans was BOLD (GE, SE, GRASE), however, CBV-weighted SS-SI-VASO has now enabled probing the laminar circuitry beyond sensory areas [82] such as the dorso-lateral pre-frontal cortex (dlPFC) [272]. In **Chapter 4**, we present CBF-weighted 3D-EPI ASL and demonstrate

[3] https://caseforge.co/

functional CBF mapping at sub-millimetre isotropic resolution for the first time in humans [207]. By showcasing the feasibility for laminar fMRI using CBF, we add yet another to the conundrum of choices for functional neuroimaging. Going forward, more studies using non-BOLD techniques addressing analysis and interpretability of the measured signals will hopefully see increased adoption to cognitive neuroscience applications taking full advantage of their quantitative nature and relatively better spatial specificity to neuronal activation compared to BOLD. Nevertheless, due to its popularity and unrivalled sensitivity, the BOLD contrast will stand the test of time. While major strides have been made in understanding the laminar specificity of the BOLD signal by modelling studies under the assumptions of steady-state [110] and more recently, from a fully dynamical [270] perspective within the framework of physiologically-informed dynamic causal modelling (P-DCM) [277], model-based deconvolution of laminar BOLD responses have only just started seeing application [195, 207] and additional validation studies will propel these model-driven techniques into mainstream adoption in laminar fMRI.