

Fourier Tract Sampling (FouTS): A framework for improved inference of white matter tracts from diffusion MRI by explicitly modelling tract volume

Thomas G. Close^{a,b,c,*}, Jacques-Donald Tournier^{d,e,f}, Leigh A. Johnston^{b,c,e}, Fernando Calamante^{d,e,f}, Iven Mareels^b, Alan Connelly^{d,e,f}

^a*Okinawa Institute of Science and Technology Graduate University, Okinawa, Japan*

^b*Department of Electrical and Electronic Engineering, University of Melbourne, Melbourne, Victoria, Australia.*

^c*National ICT Australia, Victorian Research Laboratory, Melbourne, Victoria, Australia.*

^d*Florey Institute of Neuroscience and Mental Health, Heidelberg, Victoria, Australia*

^e*Florey Department of Neuroscience and Mental Health, University of Melbourne, Melbourne, Victoria, Australia*

^f*Department of Medicine, Austin Health and Northern Health, University of Melbourne, Victoria, Australia*

Abstract

Diffusion MRI tractography algorithm development is increasingly moving towards global techniques to incorporate “downstream” information and conditional probabilities between neighbouring tracts. Such approaches also enable white matter to be represented more tangibly than the abstract lines generated by the most common approaches to fibre tracking. However, previously proposed algorithms still use fibre-like models of white matter corresponding to thin strands of white matter tracts rather than the tracts themselves, and therefore require many components for accurate representations, which leads to poorly constrained inverse problems. We propose a novel tract-based model of white matter, the ‘Fourier tract’, which is able to represent rich tract shapes with a relatively low number of parameters, and explicitly decouples the spatial extent of the modelled tract from its ‘apparent connection strength (ACS)’. The Fourier tract model is placed within a novel Bayesian framework, which relates the tract parameters directly to the observed signal, enabling a wide range of acquisition schemes to be used. The posterior distribution of the Bayesian framework is characterised via Markov-chain Monte-Carlo sampling to infer probable values of the ACS and spatial extent of the imaged white matter tracts, providing measures that can be directly applied to many research and clinical studies. The robustness of the proposed tractography algorithm is demonstrated on simulated basic tract configurations, such as curving, twisting, crossing and kissing tracts, and sections of more complex numerical phantoms. As an illustration of the approach *in vivo*, fibre tracking is performed on a central section of the brain in three subjects from 60 direction HARDI datasets.

Introduction

Due to the barriers white matter (WM) axons present to molecular motion, WM tracts can be inferred from MRI datasets by sensitising the image intensities to the diffusion of water, a technique known as diffusion MRI (dMRI) tractography (Johansen-Berg and Behrens, 2009; Jones, 2011). As the only non-invasive technique for delineating WM tracts, dMRI tractography has found a wide range of applications, from longitudinal studies of WM degeneration (Johansen-Berg and Behrens, 2006; Ciccarelli et al., 2008) to informing surgical planning in order to avoid resecting eloquent tracts (Nimsky et al., 2006; Bello et al., 2008; Farquharson et al., 2013) and studies of network connectivity in the human brain (Bullmore and Sporns, 2009; Gong et al., 2009). Depending on the application, the characteristics of WM tracts that are of interest can vary significantly. For example, determining the spatial extent of critical WM tracts should reduce the risks of serious post-surgical deficits (Nimsky et al., 2006; Bello et al., 2008; Farquharson et al., 2013), while the number of constituent axons in a WM tract, its ‘connection strength’, is of interest in studies of WM degeneration and graph theoretical analysis of WM networks (Johansen-Berg and Behrens, 2006; Ciccarelli et al., 2008; Bullmore and Sporns, 2009). However, such distinctions are often not explicit in the design of tractography algorithms, leading to the development of WM representations that cannot be directly related to any physical characteristics of WM tracts (Jones, 2010).

Despite currently being the most popular approach to dMRI tractography, streamlines fibre tracking (Mori et al., 1999; Conturo et al., 1999; Behrens et al., 2003) suffers from a number of limitations that stem from its formulation as a tracking problem (Kreher et al., 2008; Jbabdi et al., 2007; Jones, 2010). Common to all streamlines fibre tracking algorithms is the computationally inexpensive idea of stepping along local estimations of fibre orientations to produce “fibre tracks”. A limitation of this approach is that downstream data does not influence the current step (Kreher et al., 2008; Jbabdi et al., 2007), leading to the accumulation of uncertainties along the track and potentially the erroneous result that the support for the existence of a pathway depends on the end it is seeded from. Another issue is that the streamlines are independent from each other and therefore the conditional probability between potential pathways is not accounted for, i.e. the same orientation data can be used to explain any number of the generated pathways (Sherbondy et al., 2009). However, a fundamental issue is

*Corresponding author

Email address: `tcclose@oist.jp` (Thomas G. Close)

that the streamlines themselves are abstract lines derived from “tangents to the space-curve” of fibre orientations, not a true estimation of physical characteristics of WM tracts (Jones, 2010).

Several methods have recently been proposed to address the various limitations of streamlines fibre tracking (Jbabdi et al., 2007; Kreher et al., 2008; Fillard et al., 2009; Sherbondy et al., 2009; Reisert et al., 2011; Smith et al., 2013b; Schreiber et al., 2014). One approach to incorporate downstream data is to use a spline model of WM to test the probability of a connection existing between two points (Jbabdi et al., 2007; Schreiber et al., 2014). By concurrently fitting the spline model to fibre orientation estimates, downstream data informs the local estimation of the WM pathway. However, due to the difficulty of optimising the spline control points, they must be initialised near to the pathway in order to ensure convergence (Schreiber et al., 2014). Furthermore, the splines are fit to the dMRI data independently in these methods, and therefore do not account for the conditional probabilities with neighbouring WM tracts.

In order to incorporate the conditional probabilities between WM pathways, the pathways must be inferred concurrently. However, only a few published methods have taken this approach, since the complexity of the problem and the computational resources required increase considerably. By reinterpreting streamlines as a model of WM representing a specific volume of axons, the BlueMatter and SIFT algorithms prune a large number of streamlines into the subset that best represents the dMRI data (Sherbondy et al., 2009; Smith et al., 2013b). This approach is a significant improvement over typical streamlines fibre tracking, not only because the conditional probabilities between paths are considered but also because the reinterpretation of the streamlines as actual WM paths allows them to be readily applied to real world problems in brain connectivity (Smith et al., 2013a). However, although the original superset of streamlines considered is typically very large, the accuracy of the representation of the volume and shape of the true WM paths provided by the optimal subset of streamlines remains dependent on the limitations of the fibre tracking algorithm used to generate them.

Another class of dMRI tractography methods that incorporates both the conditional probabilities between WM paths and downstream data takes the approach of annealing small WM fragments into chains representing WM paths (Kreher et al., 2008; Fillard et al., 2009; Reisert et al., 2011). This is typically performed by optimising their position and orientation with respect to a cost function that penalises misalignment with the observed dMRI data and tract curvature, and terms to encourage the fragments to connect to each other into longer chains. Like the streamline pruning algorithms, the paths generated by these algorithms can be interpreted as a direct representation of WM tracts. A weakness of this approach is that the inferred WM paths are sensitive to the heuristically tuned weights of the curvature and linking

terms, which are required to constrain the inverse problem due to the high number of fragments required to accurately represent the WM structures. Furthermore, the chains formed by annealing approaches do not directly model the connection strength or spatial extent of WM tracts. In order to accurately estimate the connection strength of a tract, high numbers of thin chains would be required, since the connection strength is represented simply by the chain count. However, increasing the number of independent chains also increases the number of parameters in the model, thereby reducing the effective signal-to-noise ratio (SNR) for each inferred parameter and leading to poorly constrained inverse problems.

In order to reformulate dMRI tractography to address the inference of connection strength and spatial extent directly, we propose a novel mathematical description of WM fibre bundles, ‘Fourier tracts’, which can represent WM bundles over a range of scales, from short tract segments to whole WM tracts, with a relatively small number of parameters. Fourier tracts explicitly define the spatial extent and the and the intra-axonal volume of the WM tracts they represent, and therefore their correspondence to tract characteristics of interest is well defined. The posterior probability distribution over these parameters given the observed dMRI dataset is inferred within a novel Bayesian framework, which relates the modelled tracts directly to the observed signal. While a closed-form expression for the posterior distribution is unlikely to exist, it can be characterised using Markov-chain Monte Carlo (MCMC) sampling, which is demonstrated in this study on various artificial and *in vivo* dMRI datasets. The resulting dMRI tractography method is able to unambiguously infer the posterior distribution over the spatial extent and the apparent connection strength (ACS) of WM tracts, and therefore the resultant distributions can be directly applied to a wide range of clinical and scientific problems.

Methods

The proposed Bayesian framework consists of three components: the Fourier tract model for representing WM tracts, a prior probability distribution over the Fourier tract parameters representing the plausibility of a given tract morphology, and an efficient yet flexible signal model based on arbitrary axially-symmetric diffusion response functions (Tournier et al., 2004). Combining the signal model with an appropriate noise model completes the expression for the likelihood function and the Bayesian framework. Metropolis-Hastings (MH) MCMC sampling is then used to characterise the posterior distributions over the Fourier tract parameters.

The Fourier tract model

Instead of representing a thin strand of WM or a streamline, Fourier tracts model a continuous volume of WM fibres that can smoothly vary in

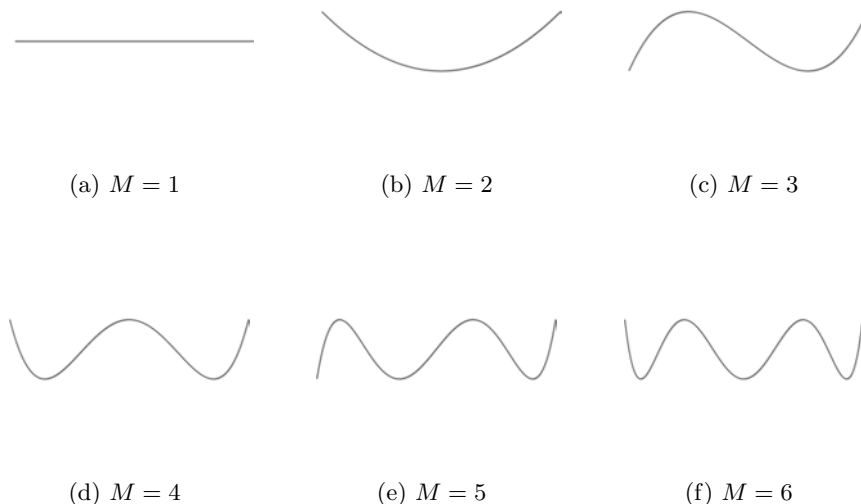


Figure 1: Example Fourier descriptors (FD) of increasing degree, M . The first order parameters of each FD, $\boldsymbol{\theta}_1$, are a horizontal vector. For each of the $M > 1$ FDs (b - f), the parameter vector, $\boldsymbol{\theta}_M$, is aligned vertically and of length $\frac{1}{4}\|\boldsymbol{\theta}_1\|$. As the degree of the FDs increase, the number of bends it can represent increases likewise. Note that the degree of a FD does not limit its maximum curvature since the parameterisation speed of the FD varies along its length. The combination of parameter vectors orientated along arbitrary axes enables the representation of a wide range of curves even for FDs of low degree.

shape over the cross-section of the tract. The volume of WM a Fourier tract represents is defined by augmenting parameters of a mean path, described by a Fourier descriptor (FD) (Chung et al., 2010), with two auxiliary FD parameter sets. The auxiliary FD parameters define how the shape and location of the constituent fibres deviate from the mean path over the cross-section of the Fourier tract.

The three-dimensional curves described by FDs have the form

$$\boldsymbol{\lambda}(\tau, \tilde{\Theta}) = \sum_{m=0}^M \boldsymbol{\theta}_m \psi_m(\tau), \quad \tau = [0, 1], \quad (1)$$

where

$$\psi_m(\tau) = \begin{cases} 1, & m = 0 \\ \sqrt{2} \cos(m\pi\tau), & m = 1, \dots, M \end{cases}, \quad (2)$$

are the basis functions, and $\tilde{\Theta} = \{\boldsymbol{\theta}_m : m = 0, \dots, M\}$ is the set of three-dimensional vectors (x , y and z axes) that comprise the FD parameterisation (Figure 1).

The Fourier tract model adds two free auxiliary parameter sets to define a two-dimensional disc embedded within the higher order FD parameter

space,

$$\begin{aligned}\gamma(\tau, \mu, \nu, \Theta) &= \lambda(\tau, \tilde{\Theta}^{(0)} + \mu\tilde{\Theta}^{(1)} + \nu\tilde{\Theta}^{(2)}), \\ \tau &= [0, 1], \mu^2 + \nu^2 \leq 1,\end{aligned}\quad (3)$$

where $\tilde{\Theta}^{(0)}$ is the parameter set of the central path, $\tilde{\Theta}^{(1)}$ and $\tilde{\Theta}^{(2)}$, the auxiliary parameter sets, and μ and ν auxiliary parameterisation variables. However, to study the relationship between these parameters and the tract shapes they correspond to, it is more convenient to decompose the FD model into a primary 3-vector, $\boldsymbol{\theta}_m^{(0)}$, and two auxiliary 3-vectors, $\boldsymbol{\theta}_m^{(1)}$ and $\boldsymbol{\theta}_m^{(2)}$ for each FD order, m .

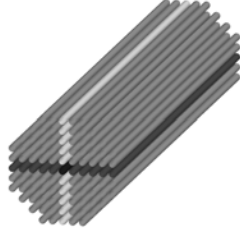
In the case of an $M = 1$ Fourier tract with

$$\begin{aligned}\boldsymbol{\theta}_1^{(0)} &\perp \boldsymbol{\theta}_0^{(1)} \perp \boldsymbol{\theta}_0^{(2)}, \\ \boldsymbol{\theta}_1^{(1)} &= \boldsymbol{\theta}_1^{(2)} = \mathbf{0}, \\ \|\boldsymbol{\theta}_0^{(1)}\| &= \|\boldsymbol{\theta}_0^{(2)}\|\end{aligned}\quad (4)$$

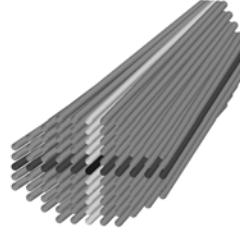
where $\boldsymbol{\theta}_0^{(0)}$, $\boldsymbol{\theta}_1^{(1)}$, $\boldsymbol{\theta}_0^{(1)}$ and $\boldsymbol{\theta}_0^{(2)}$ are the zeroth and first degree components of the primary Fourier descriptor, and zeroth degree components of the two auxiliary axes respectively. The corresponding fibres take the form of a cylindrical bundle with radius $\|\boldsymbol{\theta}_0^{(1)}\| = \|\boldsymbol{\theta}_0^{(2)}\|$ (Figures 2a and 3a). By adding non-zero first and second order terms to the auxiliary parameters, the constituent fibres can be made to fan out (Figures 2b and 3b), bunch together (Figures 2c and 3c), twist (Figures 2d and 3d), or assume more complex configurations like those expected *in vivo* through combinations of these characteristics via arbitrary FD paths. The number of potential bends in a Fourier tract increases with its degree, enabling complete WM tracts to be represented by single Fourier tracts.

Unlike previous WM models, the number of fibres a Fourier tract represents is considered independently from its spatial extent. Although variations in axonal diameter will affect the quantity of axons contained within a given cross-sectional area, it is assumed that the total intra-axonal volume stays relatively constant along the length of a tract, at least to the extent that its contribution to the DW signal is not significantly affected. Given the assumption that number of fibres in the tract is constant, any variations in the density of a non-branching tract can be explained solely by variations in the cross-sectional area of the tract. This constant ‘‘fibre count’’ is included in the Fourier tract model as the parameter, η , which we shall also refer to as the ‘‘apparent connection strength’’ (ACS) of the tract.

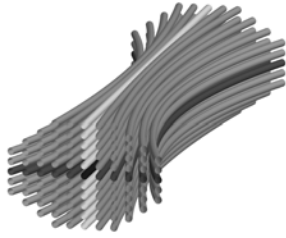
In theory, η could be directly incorporated into the Fourier tract model. However, when using iterative methods to explore the parameter space, such as MCMC sampling, it is convenient to define η indirectly to constrain the



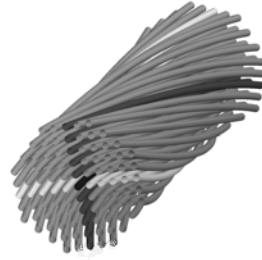
(a) $\theta_m^{(l)} = 0, l = 1, 2, m = 1, 2$



(b) $\theta_{1,y}^{(1)} > 0$



(c) $\theta_{2,y}^{(1)} > 0$



(d) $\tilde{\Theta}^{(1)}$ & $\tilde{\Theta}^{(2)}$ form arcs $\perp \tilde{\Theta}^{(0)}$

Figure 2: Example $M=2$ Fourier tracts with $\theta_1^{(0)} \perp \theta_0^{(1)} \perp \theta_0^{(2)}$ and zero higher order auxiliary parameters (i.e. $\{\theta_m^{(l)}, l = 1, 2, m = 1, 2\}$) unless otherwise specified. Each Fourier tract is represented by a grid of FDs (drawn as tubes), where the primary FD is shown in black, and the first and second auxiliary axes are shown in darker and lighter shades of grey respectively. (a) A straight Fourier tract (zero higher order parameters). (b) A Fourier tract with a non-zero $\theta_1^{(1)}$ in the same direction as $\theta_0^{(1)}$, resulting in a ‘fanning’ tract. (c) A Fourier tract with a non-zero $\theta_2^{(1)}$ aligned the y -axis in the opposite direction as $\theta_0^{(1)}$, resulting in a ‘bunching’ tract. (d) A Fourier tract with arc-shaped $\tilde{\Theta}^{(1)}$ and $\tilde{\Theta}^{(2)}$ FDs in the plane of $\theta_0^{(1)}$ and $\theta_0^{(2)}$, resulting in a ‘twisting’ tract.

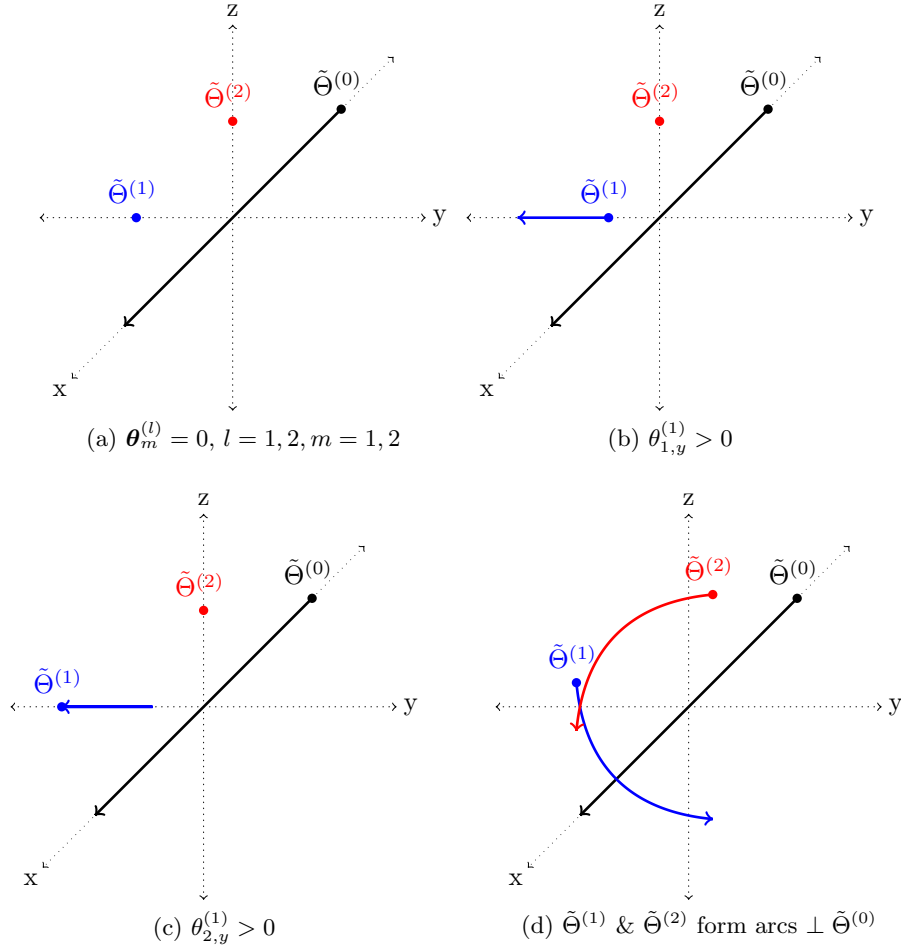


Figure 3: The primary and auxiliary FDs of the example Fourier tracts in Figure 2. Primary FDs are shown in black, and the 1st and 2nd auxiliary FDs in blue and red respectively. The FDs begin at the filled circles and travel in the direction of the arrows w.r.t τ . (a-d) correspond to the example tracts shown in 2(a-d) respectively. (a) The auxiliary FDs have no length and are located perpendicular to the primary FD. (b) The first auxiliary FD moves away from the origin in a perpendicular direction to the primary FD. (c) The first auxiliary FD moves towards the origin in a perpendicular direction to the primary FD then doubles back on itself to return to its starting location. (d) The two auxiliary FDs form arcs in the plane perpendicular to the primary FD.

values it can assume. The most important constraint is to ensure η is non-negative, which is ensured by replacing η in the model by a supporting parameter α such that

$$\eta(\Theta, \alpha) \geq \alpha^2 . \quad (5)$$

An additional tunable term, $\zeta(\Theta)$, is included into the definition of η in order to avoid configurations that contain long and thin or ribbon-like Fourier tracts with extremely low ACS values. Since Fourier tracts with extremely low ACS contribute little to the expected signal, the shape of such Fourier tracts has little impact on the likelihood function and can be safely ignored. Unlike the non-negative constraint, preventing thin and ribbon-like tracts with extremely low ACS is not strictly necessary for the integrity of the framework. However, such configurations correspond to highly correlated ridges in the prior distribution landscape, which can drastically degrade the speed of convergence of MCMC sampling algorithms to high probability regions (Gilks et al., 1996). Therefore, we chose to define η as

$$\eta(\Theta, \alpha) = \alpha^2 + \zeta(\Theta) , \quad (6)$$

where

$$\zeta(\Theta) = \epsilon_l \sqrt{\|\boldsymbol{\theta}_1^{(0)}\|} + \epsilon_w \left(\|\boldsymbol{\theta}_0^{(1)}\| + \|\boldsymbol{\theta}_0^{(2)}\| \right) , \quad (7)$$

and ϵ_l and ϵ_w are empirically tuned to set a minimum ACS that depends on the first degree length and width terms of the Fourier tract (Appendix A). Note that since α only enters the Bayesian framework in (6), and is otherwise unconstrained, indirectly expressing the ACS in this manner does not affect its inference away from the lower bound of $\zeta(\Theta)$.

Prior distribution

The expression for the prior probability of a Fourier tract configuration is designed to incorporate prior knowledge about the physical properties of WM tracts, such as their curvature and density, in order to effectively restrict the Fourier tracts to realistic configurations during the initial transient of the MCMC algorithm. Due to the lack of quantitative data for the physical characteristics of WM tracts, the prior is empirically tuned so that its contribution to the variation in the posterior distribution is low compared to the likelihood function for any set of reasonable tract configurations.

Specifically, the prior addresses four types of tract configurations that are assumed to be improbable: tracts with large high frequency oscillations, tracts that double back on themselves, tracts with implausibly high densities and ‘‘tracts’’ with densities that are so low that they would not typically be considered a coherent tract (Figure 4). These assumptions are captured in three independent terms: a ‘frequency prior’, P_f , a ‘hook prior’, P_h , and a

‘density prior’, P_d , which are multiplied to give the expression for the prior probability of a single tract,

$$\log[P(\Theta, \alpha)] = \log[P_f(\Theta)] + \log[P_h(\Theta)] + \log[P_d(\Theta, \alpha)]. \quad (8)$$

The complete prior expression for a Fourier tract configuration consisting of K tracts defined by the parameter set $\Theta = \{\Theta_k, \alpha_k : k = 1, \dots, K\}$, is then the sum of the prior probabilities of each tract

$$\log[P(\Theta)] = \sum_{k=1}^K \log[P(\Theta_k, \alpha_k)]. \quad (9)$$

The frequency prior is derived from the observation that FDs fit to typical WM paths exhibit an exponential decay with increasing degree (Chung et al., 2010). This prior is loosely analogous to the curvature constraints and cost function components used in other tractography techniques (Johansen-Berg and Behrens, 2009; Jones, 2011), with the exception that its global definition allows more flexibility in local curvature (Appendix B.1). The hook prior is derived from the assumption that tracts are unlikely to double back on themselves and is heuristically defined by comparing the displacement from samples along the tract to each endpoint with the displacement between the endpoints themselves (Appendix B.2). The density prior not only ensures that the ACS values are within a plausible range given the width of the tract, but also prevents physically impossible configurations like the one shown in Figure 4c where the Fourier tract turns itself inside out, since at the point of inflexion the density of the tract approaches infinity (Appendix B.3).

Signal model

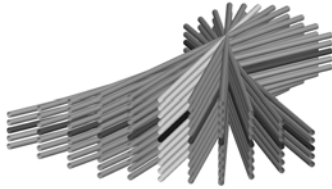
In the proposed signal model, the molecular diffusion of water is assumed to operate in a slow exchange regime with respect to both the radius of curvature of the WM paths and between distinct fibre bundles on the time-scale of the dMRI experiment (Wedeen et al., 2005). With this assumption, the dMRI signal arising from a Fourier tract can be decomposed into the sum of the signals arising from linearised segments of all Fourier tract paths in the configuration. The signal observed at a voxel arising from a segment is the product of the diffusion weighting function, $W(\mathbf{u})$, and the point-spread function between the location of the segment and the centre of the voxel, $H(\mathbf{p})$ (Tournier et al., 2004; Neeman et al., 1991) (Figure 5). The signal in the n^{th} voxel for the i^{th} diffusion encoding arising from a Fourier tract defined by parameters, $\{\Theta, \alpha\}$, is the product of the diffusion response and point-spread functions integrated over the spatial extent of the tract and



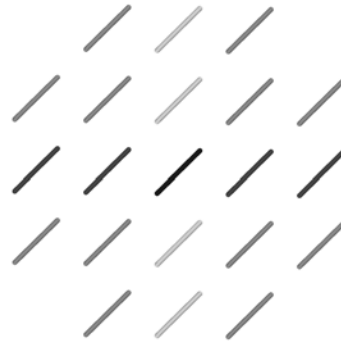
(a) $\boldsymbol{\theta}_2 \gg \boldsymbol{\theta}_1$ & $\boldsymbol{\theta}_2 \perp \boldsymbol{\theta}_1$



(b) $\|\boldsymbol{\theta}_2\| \cdot \|\boldsymbol{\theta}_1\| \approx 1$



(c) $\|\boldsymbol{\theta}_0^{(1)}\| < \|\boldsymbol{\theta}_1^{(1)}\|$



(d) Large $\|\boldsymbol{\theta}_0^{(1)}\|$ & $\|\boldsymbol{\theta}_0^{(2)}\|$ & low η

Figure 4: Configurations rendered improbable by the prior distribution. (a) A FD with $\|\boldsymbol{\theta}_2\| > \|\boldsymbol{\theta}_1\|$ forming a sharply curving arc. (b) A FD with $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ nearly colinear, which forms a sharp hook. (c) A Fourier tract with $\|\boldsymbol{\theta}_1^{(1)}\|$ greater than $\|\boldsymbol{\theta}_0^{(1)}\|$, resulting in an infinitely dense line of inflection which the entire cross-section passes through. (d) A wide Fourier tract with a low density of fibres.

scaled by the ACS of the tract,

$$\hat{s}_{i,n}(\Theta, \alpha) = \eta(\Theta, \alpha) \int_{-1}^1 \int_{-a(\nu)}^{a(\nu)} \int_0^1 W_i \left(\frac{\gamma'(\tau, \mu, \nu, \Theta)}{\|\gamma'(\tau, \mu, \nu, \Theta)\|} \right) H_{i,n}(\gamma(\tau, \mu, \nu, \Theta)) \|\gamma'(\tau, \mu, \nu, \Theta)\| d\tau d\mu d\nu \quad (10)$$

where $a(\nu) = \sqrt{1 - \nu^2}$ and $\gamma'(\tau, \mu, \nu, \Theta) = \frac{d\gamma(\tau, \mu, \nu, \Theta)}{d\tau}$, which like $\gamma(\tau, \mu, \nu, \Theta)$ is a linear transform of Θ .

Note that since (10) is not normalised by $\left\| \frac{d\gamma(\tau, \mu, \nu, \Theta)}{d\mu} \right\|$ or $\left\| \frac{d\gamma(\tau, \mu, \nu, \Theta)}{d\nu} \right\|$ the number of modelled fibres stays constant as the cross-sectional area of a Fourier tract varies along its length; locations where a Fourier tract widens represent a reduction of the tract's density, which can occur *in vivo* when interdigitating with another tract or spreading into grey matter. The complete expression for the signal in the n^{th} voxel for the i^{th} encoding given the parameters for a Fourier tract configuration, Θ , is the sum of individual tract contributions over all Fourier tracts in the configuration,

$$\hat{S}_{i,n}(\Theta) = \sum_{k=1}^K \hat{s}_{i,n}(\Theta_k, \alpha_k). \quad (11)$$

An analytical expression for (10) is unlikely to exist for any reasonable choice of $W(\mathbf{u})$ and $H(\mathbf{p})$. Therefore, (10) is numerically approximated by the sum of short segments at T_s regularly spaced samples along the primary axis and U_s regularly spaced samples along both the auxiliary axes such that $\mu^2 + \nu^2 \leq 1$. T_s and U_s are empirically chosen to balance the accuracy of the approximation with computational performance (Appendix A).

Unless otherwise specified, the diffusion response function used throughout this manuscript is derived from an assumed diffusion tensor model (Basser et al., 1994) with a fractional anisotropy of 0.8 and a mean diffusivity of $5 \times 10^{-4} \text{ mm}^2\text{s}^{-1}$ for a 60 direction $b = 3000 \text{ s/mm}^2$ encoding scheme. For the imaging acquisition, a typical 2D Echo-Planar imaging (EPI) scheme with an apodized slice selection pulse is assumed. The point spread function is modelled by a truncated sinc function along the frequency and phase encoding axes of a echo-planar imaging (EPI) acquisition, and a polynomial function approximates the apodized slice selection profile along the slice selection axis (Appendix C). However, any arbitrary diffusion response and point-spread functions can be substituted into (10) to match the dMRI acquisition scheme used.

Noise model

Noise in MR magnitude images has been shown to have a Rician distribution (Gudbjartsson and Patz, 1995). However, it is convenient to approximate the Rician distribution with a Gaussian so that the log likelihood is

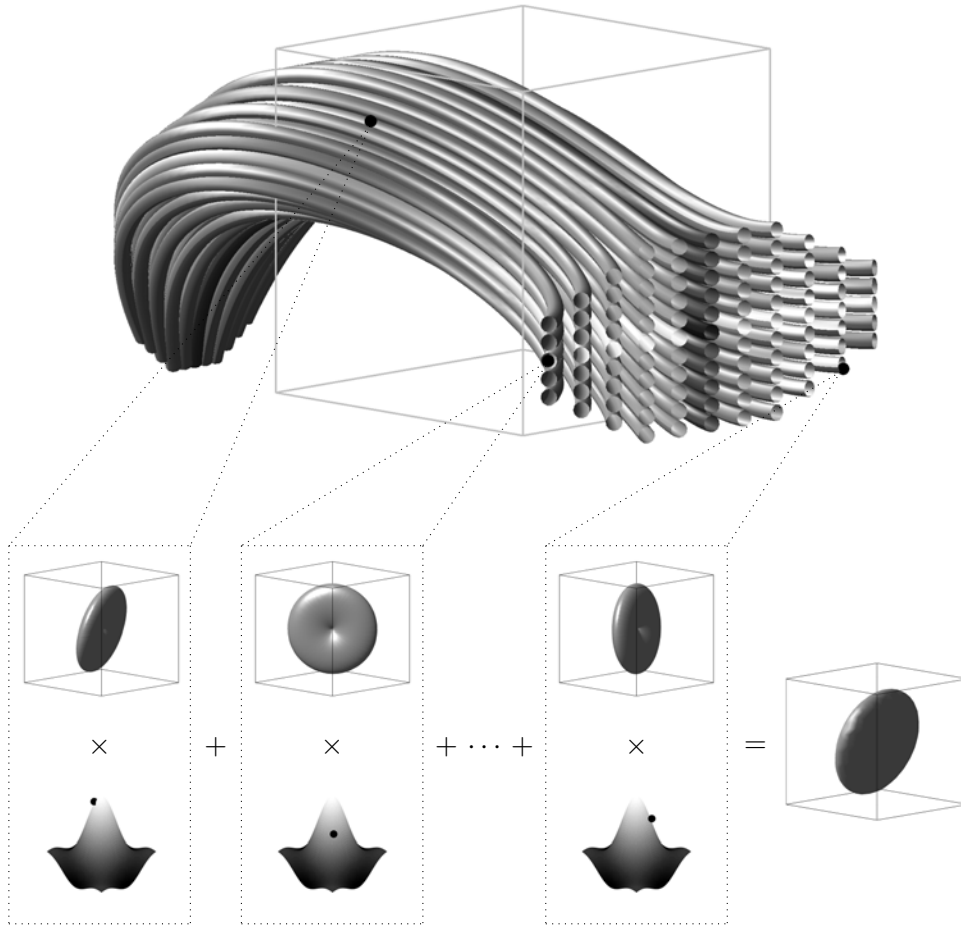


Figure 5: Illustration of the method for modelling the contribution of a Fourier tract to the signal of a voxel. (*Top*) A Fourier tract passing through the outline of a voxel. The insets show the contributions of three example segments to the modelled signal. (*Insets-top*) Diffusion response function aligned to the segment. The displacement from the origin to the surface represents the expected signal for $b=3000$ s/mm² gradients aligned to the corresponding orientation. (*Insets-bottom*) truncated sinc point-spread function centred on the voxel within the imaging plane. Black circles represent the locations of the segments on the point-spread function landscape relative to the voxel. (*Bottom-right*) the modelled signal of the voxel: the sum of the weighted diffusion response functions arising from all points within the Fourier tract.

polynomial. Although this is a poor approximation when signal-to-noise is lower than 2, which commonly occurs in directions of high diffusion weighting, idealised simulations of a 60-directional $b = 3000$ diffusion weighting scheme show that when averaged over all directions, a Gaussian noise model is good approximation for SNRs in the $b = 0$ image greater than 10 (see supplementary material S.1). In this case, the expression for the log likelihood of the set of observed signals, $\mathbf{S} = \{S_{i,n} : i = 1, \dots, I, n = 1, \dots, N\}$, given a Fourier tract configuration parameter set, Θ , is

$$\log[\mathbf{P}(\mathbf{S}|\Theta)] = \log \left[\frac{\sigma\sqrt{2\pi}}{NI} \right] - \frac{1}{\sigma^2} \sum_{n=1}^N \sum_{i=1}^I (S_{i,n} - \hat{S}_{i,n}(\Theta))^2, \quad (12)$$

where σ is the variance of the assumed Gaussian noise.

To avoid issues caused by partial voluming with isotropic signal arising from surrounding grey matter and cerebrospinal fluid (Alexander et al., 2001), the isotropic components of both the observed and modelled signals are subtracted, leaving only the non-isotropic components to be compared,

$$\log[\mathbf{P}(\mathbf{S}|\Theta)] = \log \left[\frac{\sigma\sqrt{2\pi}}{NI} \right] - \frac{1}{\sigma^2} \sum_{n=1}^N \sum_{i=1}^I \left(S_{i,n} - \frac{1}{I} \sum_{j=1}^I S_{j,n} - \hat{S}_{i,n}(\Theta) + \frac{1}{I} \sum_{j=1}^I \hat{S}_{j,n}(\Theta) \right)^2 \quad (13)$$

Note that the removal of the isotropic components from both the observed and expected data sets results in the comparison of negative signals, but that this does not affect the Likelihood since the assumed Gaussian noise is symmetric about zero.

Bayesian Formulation

The dMRI tractography problem formulated in a Bayesian framework is

$$\mathbf{P}(\Theta|\mathbf{S}) = \frac{\mathbf{P}(\mathbf{S}|\Theta)\mathbf{P}(\Theta)}{\mathbf{P}(\mathbf{S})}. \quad (14)$$

However, as is often the case in complex Bayesian frameworks, the expression for the probability of observing the given experimental data, also known as the model evidence,

$$\mathbf{P}(\mathbf{S}) = \int \mathbf{P}(\mathbf{S}|\Theta)\mathbf{P}(\Theta)d\Theta, \quad (15)$$

is analytically intractable, and therefore the posterior distribution is numerically integrated from the “unnormalised” expression for the posterior distribution,

$$\log [\tilde{\mathbf{P}}(\Theta|\mathbf{S})] = \log [\mathbf{P}(\mathbf{S}|\Theta)] + \log [\mathbf{P}(\Theta)], \quad (16)$$

noting that $\mathbf{P}(\mathbf{S})$ is a constant, normalising factor for a given dataset.

Data acquisition and preprocessing

dMRI datasets were acquired from three healthy subjects (1 right-handed male, 1 left-handed male and 1 right-handed female) using a twice-refocused spin-echo echo planar imaging sequence on a 3T Siemens Trio system: a 60 direction encoding scheme ($b=3000$ s/mm²) interleaved with 8 $b=0$ s/mm² images at $2.5 \times 2.5 \times 2.5$ mm resolution with TE=110 ms and TR=8.4 seconds. To constrain the sample locations and reduce the computational time required for the MCMC sampling, each image was trimmed to a manually selected region, which extended approximately 10~15 mm of the lateral ventricles in all directions.

In order to compare the acquired datasets with those generated by the proposed signal model, the signal-intensity-to-ACS-ratio (SIAR) was estimated from the acquired datasets. Constrained spherical deconvolution (Tournier et al., 2007) was performed within whole-brain masks (Tournier et al., 2012), and the 1% of voxels with the highest magnitude peaks in the resulting Fibre Orientation Distributions (FOD) were defined to be of unit density. A straight Fourier tract of unit density, spanning a $3 \times 3 \times 3$ voxel neighbourhood was positioned at the centre of each of these voxels, and aligned to the peak FOD direction. The preliminary SIAR estimate for each voxel was the ratio between the observed intensities and intensities simulated for the unit density tract using the proposed signal model averaged over the encoding directions. The SIAR estimate for the complete dataset was the median of the preliminary voxel-wise estimates. The assumed SNR of the likelihood function was determined independently for each voxel by the unbiased estimate of the variance in the $b=0$ images.

Results

Assessment of fibre width resolution

A possible alternative to the Fourier tract model would be one consisting of a large number of thin fibres, each with fixed signal intensities. To demonstrate the limitations of such WM models and motivate our use of the Fourier tract model, simulations were performed over a $3 \times 3 \times 3$ voxel neighbourhood to infer the configuration of increasingly larger sets of WM paths with progressively smaller contributions to the overall dMRI signal. In this case, the Fourier tract models are replaced by FDs with fixed ACS values. No noise was applied to the simulated images but an SNR of 20 was assumed for the likelihood function. Four test sets of parallel FDs were derived by sampling the auxiliary axes of a single, straight Fourier tract aligned with the first imaging axis on a square grid. The test sets contained 1, 5, 21 and 61 FDs respectively, with the ACS of each of the FDs scaled so that the total ACS of the set is 1.

For each test, one hundred samples were drawn from the posterior distribution of FD sets ($M=2$) with the same number of FDs as the reference

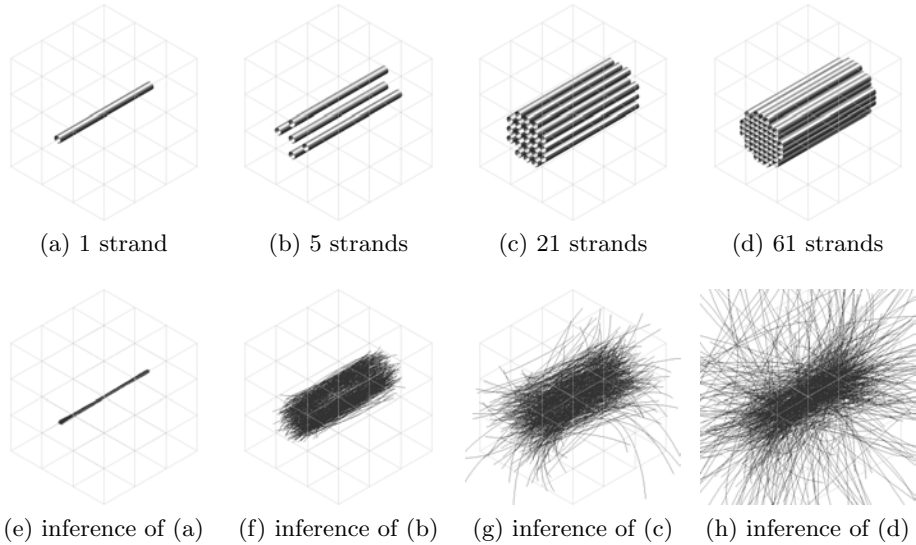


Figure 6: Metropolis Hastings Markov chain Monte Carlo samples of increasing numbers of parallel FDs. The path of each FD in each sample is represented by a single line. Samples are superimposed to represent the posterior distribution: dense populations imply probable configurations. As the number of parallel FDs is increased concurrently in both the true (a-d) and the sample (e-h) configurations from 1, 5, 21, to 61, the samples become less consistent with the true configurations. The sample period between displayed samples is increased with the number of FDs in each sample so that total number of displayed strands is approximately constant between cases ($\sim 5,000$) and the number of stray FDs can be directly compared (i.e. 1 strand case is sampled every 200 iterations, 5 strands=1,000 iterations, 21 strands=4,000 iterations, 61 strands=12,000 iterations). The increasing uncertainty from (e) to (h) demonstrates one of the pitfalls in fitting large numbers of independent WM fibre models to a small neighbourhood: the problem becomes increasingly ill-posed.

set by sampling every ten thousand iterations of a MH-MCMC chain. The MCMC chains were initialised by perturbing the parameters of the reference set by values drawn from a normal distribution, $\mathcal{N}(0, 5 \times 10^{-3})$ voxel lengths (VL). Each sample was compared with the reference configuration using the minimal linear assignment between the FDs in the sample and reference configuration, and the standard deviations of the displacement distance, $\|\boldsymbol{\theta}_{\text{sample},0} - \boldsymbol{\theta}_{\text{ref},0}\|$, angular deviation, $\boldsymbol{\theta}_{\text{sample},1} \angle \boldsymbol{\theta}_{\text{ref},1}$ and curvature magnitude, $\|\boldsymbol{\theta}_{\text{sample},2}\|$ were determined for each set of samples (Burkard et al., 2009).

The MH-MCMC generated samples inferred from the simulated datasets of the 1, 5, 21 and 61 independent FDs test cases show the trade off between accurate representation of the connection strength and accurate inference of the spatial extent of the fibre tract when the ACS is fixed per component and the number of paths alone represents connection strength (Figure 6).

The posterior distribution samples for the single FD case are tightly clus-

tered around the true FD configuration used to generate the test dataset (Figure 6e), with displacement, angular and curvature standard deviations of 0.089 VL, 0.006 rad and 0.089 VL respectively. In the 5 FD case, the samples are still largely contained within the volume of the true configuration with a displacement standard deviation of 0.121 VL, but considerably less coherently oriented than the single FD case with an angular standard deviation of 0.05 rad (Figure 6f). The inferred samples of the 21 FD test configuration contain a number of stray FDs that are no longer contained within the convex hull of the true bundle and have a displacement standard deviation of 0.21 VL and an angular standard deviation is 0.157 rad (Figure 6g). For the 61 FD case, a significant fraction of the inferred samples stray into orientations far from the true bundle orientation with an angular standard deviation of 0.36 rad and a curvature standard deviation of 0.274 VL (Figure 6h).

Assessment of inference of basic tract configurations

Eight basic test cases were chosen to examine the types of tract configurations that can be reliably inferred using the proposed framework and the convergence properties of the MH-MCMC algorithm over the posterior probability landscape:

- a tract with width less than one voxel
- a tract with width greater than one voxel
- a curving tract
- a tract with a 120° twist
- two tracts crossing at 45° to each other
- two tracts crossing at 22.5° to each other
- two tracts crossing exclusively by passing over/under each other
- two ‘kissing’ tracts.

For each of the test cases, dMRI datasets consisting of $3 \times 3 \times 3$ isotropic voxels were simulated with b values of 3000 s/mm² for 60 DW directions and one $b=0$ image. Although the absolute dimensions of the voxels are arbitrary and only the scale relative to the test configurations is relevant, the voxels are assigned an isotropic dimension of 2mm for ease of reference when quoting the inferred ACS values. The ACS of the reference Fourier tracts were set such that the average density was one. Gaussian noise was added to the simulated images such that the SNR between the maximum intensity in the $b=0$ image and noise standard deviation was 20, and a Gaussian noise model with SNR 20 was assumed in the likelihood function.

Five consecutive MH-MCMC chains (i.e. five restarts) were generated for each test configuration. Two samples were taken from each chain at an interval of 25,000 iterations after a burn-in of 75,000 iterations. Each chain was initialised with the same number of $M=2$ Fourier tracts as the corresponding test configuration. Their initial values of $\theta_1^{(0)}$, $\theta_0^{(1)}$ and $\theta_0^{(2)}$ vectors were drawn from normal distributions of $\mathcal{N}(0, \frac{4}{15})$, $\mathcal{N}(0, \frac{1}{15})$ and $\mathcal{N}(0, \frac{1}{15})$ mm respectively. The centre of masses, $\theta_0^{(0)}$, were initialised from a uniform random distribution across the bounds of the image.

In order to prevent the generated Fourier tract samples from straying outside of the bounds of the test datasets, a layer of empty voxels bounded the $3 \times 3 \times 3$ neighbourhood (analogous to embedding in isotropic medium).

Figure 7 shows the set of test configurations and the corresponding posterior distributions, which are characterised via clusters in the MCMC-generated samples. The probability distribution visualisations consist of two samples (the 75,000th and 100,000th iterations) taken from each of the 10 MCMC chains (restarts), which are overlaid on the outline of the $3 \times 3 \times 3$ voxel grid.

With the exception of the 22.5° crossing and kissing configurations, the MCMC chains consistently converged to the peak in the posterior distributions that includes the reference configuration by the 75,000th iteration. In the case of the 22.5° crossing and kissing configurations, the samples span the reference configuration and other qualitatively different configurations, such as a single straight tract and a crossing tracts for the 22.5° crossing and kissing configurations respectively. On average, the standard deviation of the likelihood function is 3.6 times larger than the standard deviation of the prior distribution in the converged MCMC chains ($>75,000$ iterations).

The cluster is noticeably broader in the case of the thinner tract configuration (Figure 7e) than the thicker configuration (Figure 7f), as represented by the tightness of the clusters. In both cases the tracts are slightly truncated lengthways, where the tract approaches the edge of the $3 \times 3 \times 3$ neighbourhood. Both the curving and twisting configuration features are correctly characterised in Figures 7g and 7h. In the case where the crossing between two tracts is 45° , two independent pathways are consistently identified. However, when the angle of incidence is 22.5° several samples correspond to a single straight bundle, representing the probability of such a configuration relative to the two independent pathways configuration. In the case of exclusive crossing, the side on which the tracts pass by each other is consistently identified within the voxel. While in the kissing case the samples are split between identifying the true kissing configuration and the nearby crossing and fanning configurations.

The box and whiskers plots of the Fourier tract ACS values show tight clusters near the true value of the test configuration (Figure 8a). For all cases there is noticeable offset of the median ACS above the true value.

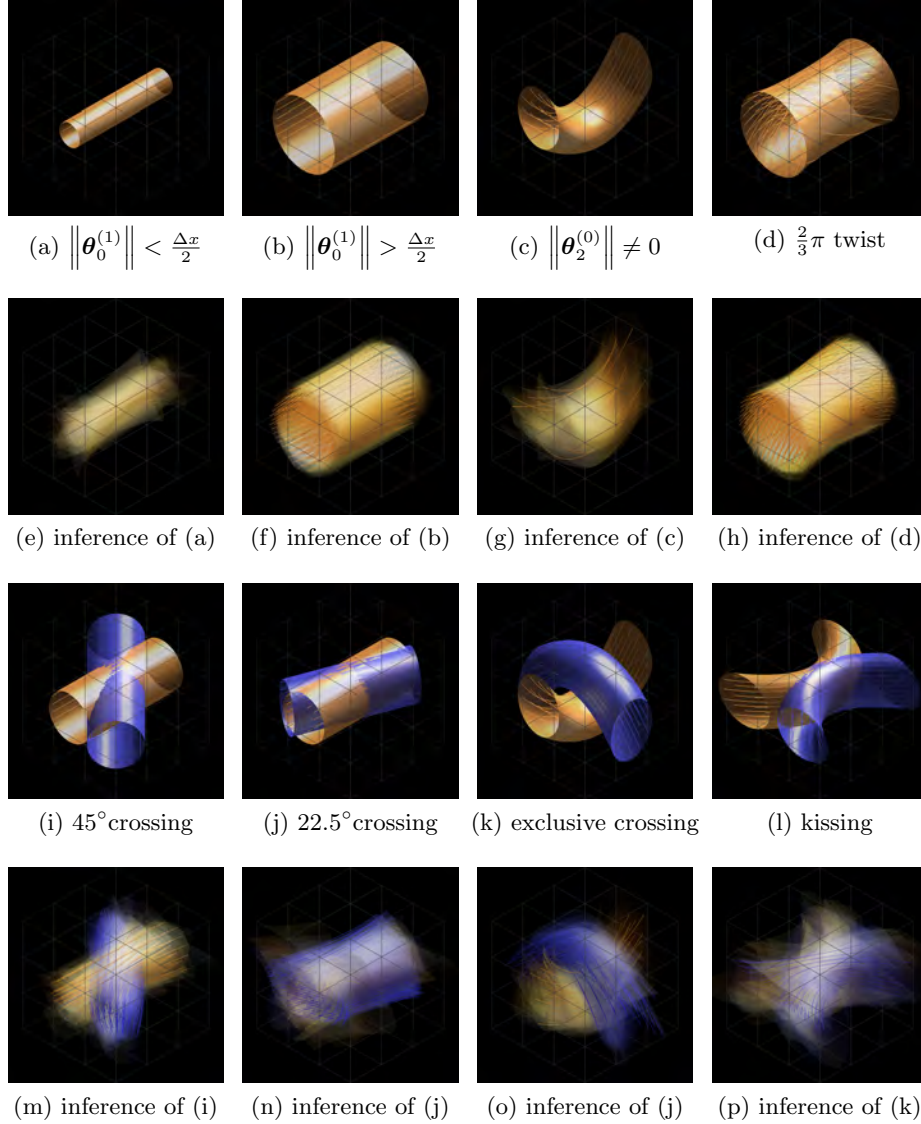


Figure 7: Testing the convergence of the sampling algorithm and the ability of the framework to resolve basic tract configurations. (a - d) Single tract reference configurations. (e - h) Samples drawn from the posterior distributions corresponding to (a - d) respectively. (i - l) Double tract reference configurations. (m - p) Samples drawn from the posterior distributions corresponding to (i - l) respectively. The inferred configurations each consist of two samples, the 75,000th and 100,000th iterations, from 10 MCMC chains overlaid on each other. The convex hulls of the test tracts and their samples are shown by semi-transparent tubes and the Apparent Connection Strength (ACS) is scaled and represented to the nearest integer by the number of solid lines sampled at fixed values μ and ν across the cross-sections of the Fourier tracts. In configurations with two tracts, the Fourier tracts in each sample are ordered, and thereby assigned colours, for the best, in a least-squares sense, match to colours assigned to the true configuration.

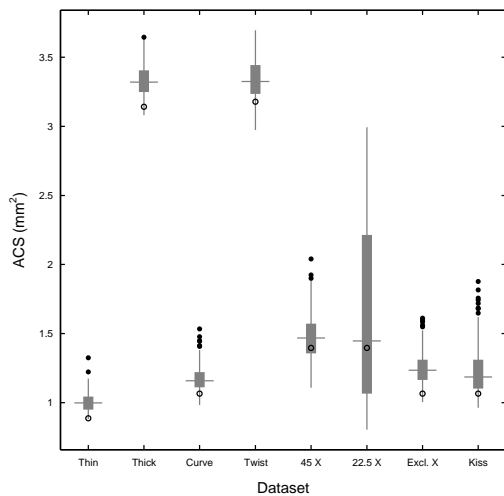
Figure 8b shows the median difference in the length of the Fourier tracts between samples and the true configuration. In all cases, there is a small underestimation of the length of the Fourier tracts in the true configuration.

Assessment of inference of intra-voxel positioning

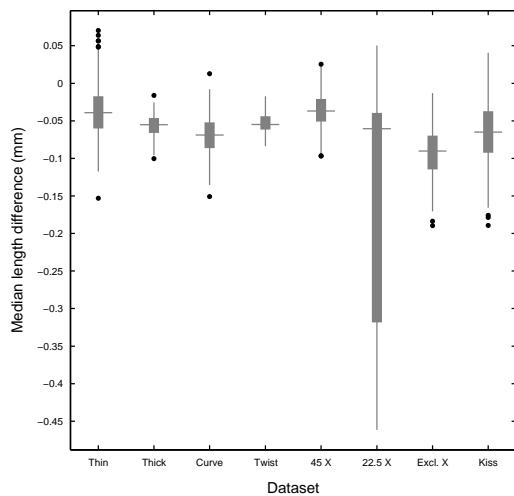
The ability of the proposed algorithm to infer the intra-voxel positioning of white matter tracts was assessed using numerical phantoms generated by the Numerical Fibre Generator tool set (Close et al., 2009). Three phantoms were selected from the default set of ten published on the Neuroimaging Informatics Tools and Resources Clearinghouse (<http://www.nitrc.org/projects/nfg/>), which were assigned an arbitrary radius length of 10mm for ease of reference. For each phantom, dMRI datasets consisting of $3 \times 3 \times 3$ 2.5mm isotropic voxel neighbourhoods centred on the phantom, were simulated using the diffusion and noise parameters used for the basic test configurations.

Five consecutive MH-MCMC chains were generated for each phantom on single cores of a 2.53 GHz Intel Xeon Processor E5649, and took 5.9 hours on average to complete. Five samples were taken from each chain at an interval of 1,000 iterations after a burn-in of 20,000 iterations. Each chain was initialised with four $M=3$ Fourier tracts with the same technique used for the basic test configurations. To constrain the Fourier tract samples to the region of interest while still allowing them to model the signal arising from fibre bundles outside the neighbourhood, “soft barriers”, 5mm along each axis from the neighbourhood centre (i.e. half a voxel width outside of the neighbourhood), were implemented via an additional ‘barrier’ prior term, P_b (Appendix B.4).

Samples from the first MCMC chain for each phantom (Figure 9 *middle column*) are shown alongside all fibre bundles (Figure 9 *right column*), and only those bundles that pass through the central voxel (Figure 9 *left column*), in the corresponding phantom neighbourhoods. The shapes, orientations and positions of the WM fibres represented by the Fourier tract samples (Figure 9 *middle column*) largely correspond to those of the reference, particularly those that pass through the central voxel of the neighbourhood (Figure 9 *left column*). However, there is not a one-to-one relationship between the spatial extents of the Fourier tract samples and those of the reference bundles, with some Fourier tracts covering the extents of two bundles (Figure 9 (b)–yellow Fourier tract, (a)–yellow and pink bundles), and others running in side-by-side within the extent of a single bundle (Figure 9 (h)–yellow and red Fourier tracts, (g)–yellow bundle). The Fourier tract samples are also influenced by the orientation and position of the surrounding fibre bundles (Figure 9 (b)–mid-blue Fourier tract, (c)–orange and pink bundles; (e)–light-blue Fourier tract, (f)–aqua bundle; (e)–mid-blue Fourier tract, (f)–purple and grey bundles; (h)–red Fourier tract, (h)–olive and mid-blue bundles).



(a) Inferred apparent connection strength (ACS)



(b) Difference length in from true configuration

Figure 8: Inference of the apparent connection strength (ACS) of basic tract configurations. (a) Box and whiskers plots of sampled ACS values for thin, thick, curving, twisting, 45° crossing, 22.5° crossing, exclusive crossing and kissing test configurations. The box, whiskers, horizontal line, solid points and hollow ring represent the interquartile range, $1.5 \times$ interquartile range extending either side of the 1st and 3rd quartiles, the median, outliers, and the ACS value of the true configuration respectively. With the exception of the 22.5° case, the sampled ACS values form tight clusters near the true values. However, the sample medians overestimate the true values in all cases. (b) Box and whiskers plot of the difference in $\|\theta_0^{(1)}\|$ between the MCMC samples and the length of the corresponding test configuration. The length of the Fourier tract samples underestimate that of the true configuration in all cases.

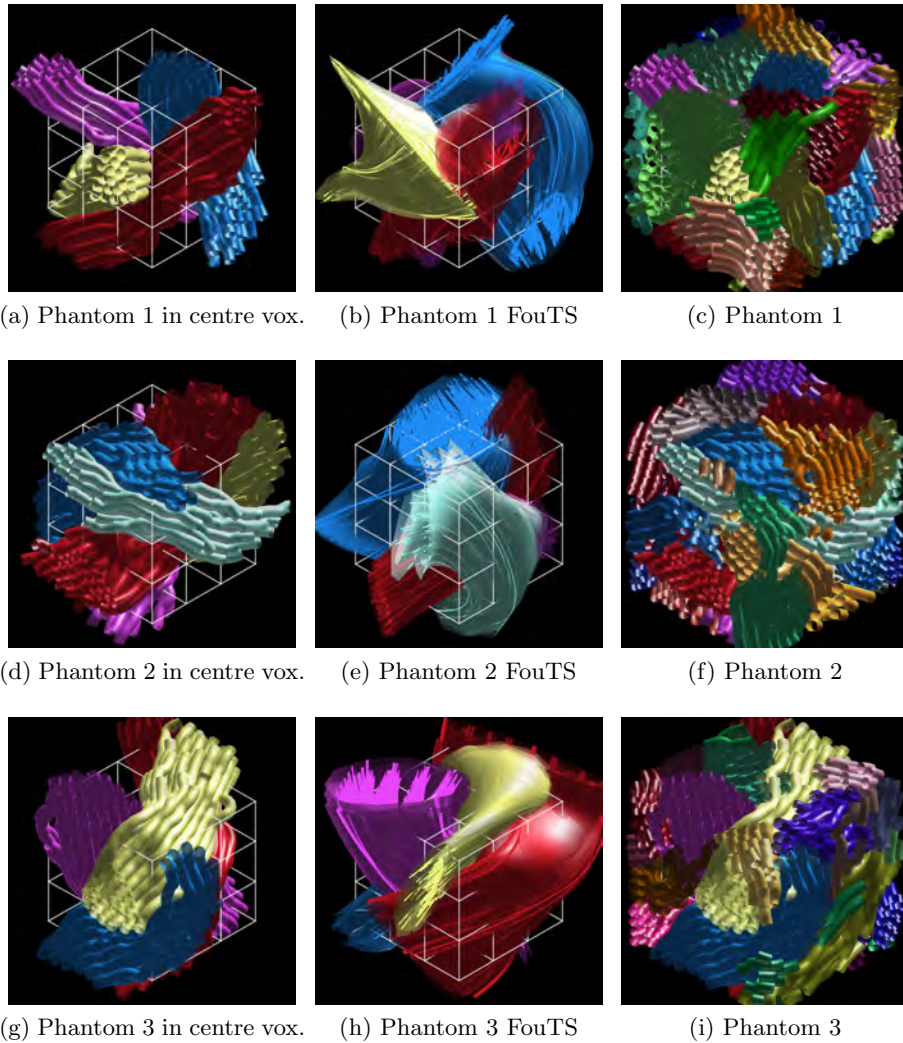


Figure 9: Inference of intra-voxel positioning and orientations via Fourier Tract Sampling (FouTS) from $3 \times 3 \times 3$ simulated images of numerical phantoms. (*Right column*) Reference bundles within the phantom neighbourhood. (*Left column*) Reference bundles that pass through the central voxel with surrounding bundles omitted from visualisation for ease of comparison. (*Middle column*) FouTS samples from the first Markov-Chain Monte-Carlo (MCMC) chain for each phantom neighbourhood. The spatial extents of the FouTS samples are represented by the semi-transparent outline, and the number of lines approximately represents the Apparent Connection Strength (ACS) of the FouTS samples. The voxel borders are shown by the white grid. Unique colours are assigned to the Fourier tracts in each sample to distinguish the tracts from each other. The assigned colours of the FouTS samples are matched to the colours assigned to the reference bundles. The locations, orientations of the constituent pathways in the Fourier tracts largely correspond to the reference bundles that pass through the central voxel (*left column*) but are influenced by the surrounding fibre bundles (*right column*).

To compare the inference of intra-voxel position and orientations of Fourier Tract Sampling (FouTS) with existing techniques, one million probabilistic streamlines (Parker et al., 2003) were generated from fibre orientation distributions estimated by Constrained Spherical Deconvolution (CSD) (Tournier et al., 2007). A maximum spherical harmonic order (l_{\max}) of 8 and a response function derived from the tensor model assumed in the simulated dataset were used for the CSD. The probabilistic streamlines were seeded throughout the complete phantom (i.e. a sphere of radius 10mm) with a step size of 0.25mm, a maximum angle of 9° and a minimum length of 5mm. Directionally-Encoded-Colour Track Density Imaging (DEC-TDI) (Calamante et al., 2010) was performed on the generated streamlines, FouTS samples and reference phantom strands with a resolution of 0.075mm in-plane and 0.44mm through-plane. The "tracks" for the FouTS samples and reference strands were generated by taking evenly-spaced samples across the Fourier tract and reference strands cross-sections (i.e. the auxiliary axes of the Fourier tracts). The number of tracks generated from each Fourier tract was approximately proportional to the ACS of the Fourier tract. The spatial extents and orientations of the reference bundles are more accurately inferred in the DEC-TDI images generated from FouTS samples than those generated from streamlines, particularly near the central voxel (Figure 10).

Assessment of in vivo tract characterisation

As a proof of concept of the proposed approach *in vivo*, the spatial extents and the ACS values of the corpus callosum and fornix stem were inferred for three healthy subjects.

The Fourier tracts were initialised from manually selected seed positions within each WM tract of interest, which were located by major anatomical landmarks in the $b=0$ images. In the case of the fornix stem, a single seed was placed in the centre of the known tract location, while in the case of the corpus callosum, nine Fourier tracts were distributed along the corpus callosum mid-line on the sagittal plane. For each MCMC chain, the initial centre of mass of the Fourier tract(s) ($M=5$), $\theta_0^{(0)}$, was perturbed from this seed point by the addition of a variable drawn from the normal distribution, $\mathcal{N}(0, 0.5)$ mm. The initial values of the $\theta_1^{(0)}$ vectors, and $\theta_0^{(1)}$ and $\theta_0^{(2)}$ vectors were drawn from random distributions $\mathcal{N}(0, 0.3)$ mm and $\mathcal{N}(0.75, 0.075)$ mm respectively, such that they typically started much smaller than a voxel in size.

For each WM tract seed point(s) (i.e. fornix and corpus callosum) of each subject (Subject 1, 2 and 3), five consecutive MH-MCMC chains were generated on a single core of a 2.53 GHz Intel Xeon Processor E5649. Two samples were taken from each chain at an interval of 25,000 iterations after a burn-in of 50,000 iterations. The average time taken to complete the 50,000 MH-MCMC iterations was 28.2 and 62.8 hours over the for the fornix and

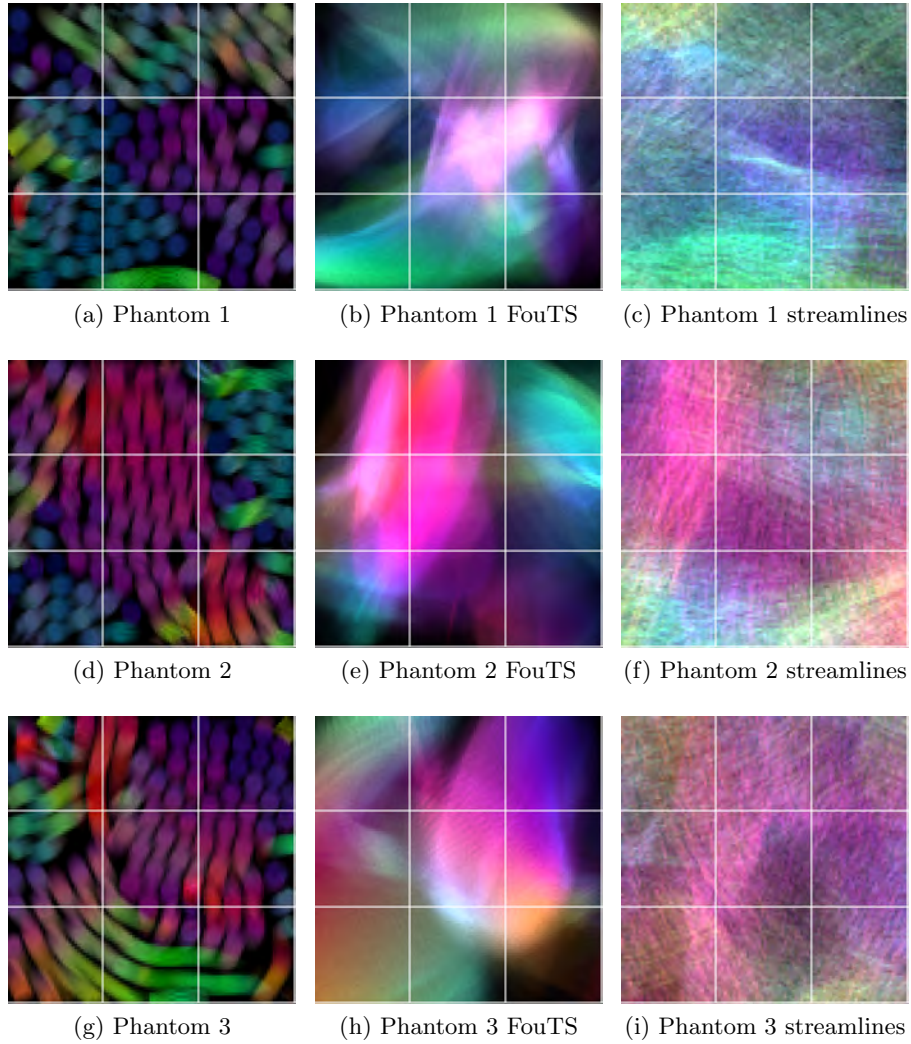


Figure 10: Directionally-Encoded-Colour Track Density Images (DEC-TDI) of fibre bundles in numerical phantoms (*left column*), Fourier Tract Sampling (FouTS) (*middle column*) and probabilistic streamlines from Fibre Orientation distributions estimated by Constrained Spherical Deconvolution (CSD) (*right column*). Red pixels indicate tracks that are aligned up-down, green left-right and blue in-out of the page. Borders of voxels in the dMRI dataset are indicated by white lines. The number of tracks generated for each Fourier tract sample was proportional to the Apparent Connection Strength (ACS) of the Fourier tract. Intra-voxel orientation and positioning of the phantom fibres is more accurately inferred by the FouTS than the probabilistic streamlines.

corpus callosum tracts respectively. Both the fornix and corpus callosum were inferred from the same region of interest for each subject, which was $28 \times 40 \times 17$, $31 \times 38 \times 13$ and $26 \times 35 \times 17$ voxels (2.5mm isotropic) in size for Subjects 1-3 respectively.

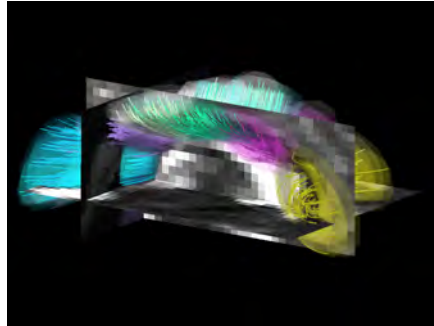
The posterior distributions over the shape and location of the corpus callosum and the fornix are characterised by samples from the generated MCMC chains (Figure 11). The 5 MCMC chains generated for each dataset robustly converge to the same peaks in the posterior distributions. The spatial extents of the corpus callosum and fornix can be inferred from the convex hulls of the visualised samples for each dataset, which overlay closely on each other with few spurious pathways. Despite the tight clustering of the Fourier tract samples over multiple initialisations, the MCMC chains did not converge to a region where mixing of the chain was observed, as slight adjustments in the Fourier tract parameters were still causing non-negligible increases in the likelihood function when the sampling was terminated due to the steepness of the peak in the posterior distribution.

Figure 12 shows the inferred ACS for the corpus callosum and fornix for each subject. The ACS distributions for the corpus callosum of the three subject overlap and have broad peaks, whereas the ACS distributions for the fornix are relatively sharp and distinct between the subjects.

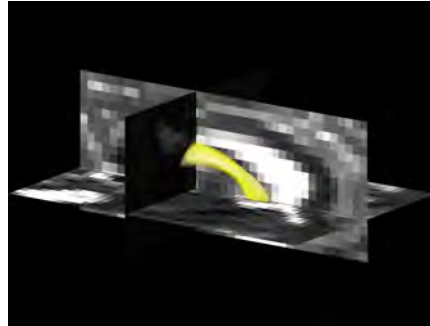
Discussion

We have proposed a novel Bayesian framework for dMRI tractography that explicitly represents the spatial extent and connection strength of WM tracts. By separating these characteristics, the proposed framework in principle should be able to directly infer quantities of practical interest, such as the risk of resecting eloquent WM in a proposed surgical operation, or whether the strength of a particular structural connection affects the probability of neurological disorders.

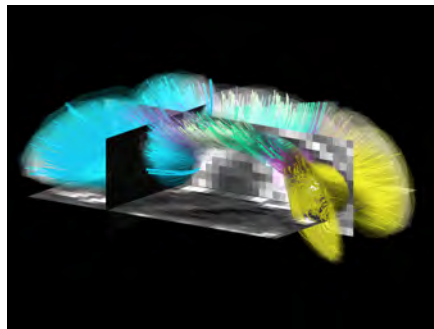
The robustness of the proposed algorithm was demonstrated by the consistent convergence of five consecutive MCMC chains to the simulated tract positions. In the 22.5° -crossing and kissing configurations, qualitatively different structures, such as a single broader tract and a crossing configuration respectively, are nearly as probable as the “true configuration” and are therefore represented by several MCMC samples. However, the inability to clearly resolve such configurations is not a limitation of the proposed algorithm, rather a limitation of the resolution of diffusion MRI itself; the configuration is therefore correctly represented as a broad peak in the posterior distribution covering the alternate structures. This effect can be seen in the merging of the peaks in the likelihood corresponding to the two “true configurations” (there are two true configurations since the order of the tracts is irrelevant) to form one long peak including the configuration in which both tracts are colinear for crossing angles below 22.5° (Figure S.2).



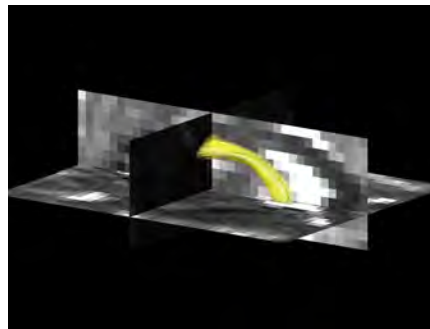
(a) Corpus callosum, Subject 1



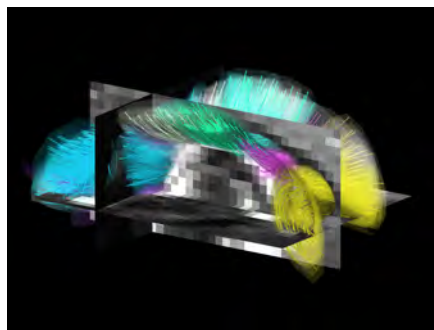
(b) Fornix, Subject 1



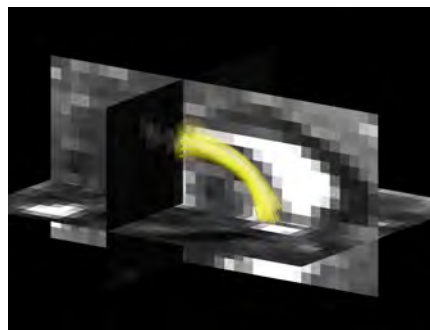
(c) Corpus callosum, Subject 2



(d) Fornix, Subject 2



(e) Corpus callosum, Subject 3



(f) Fornix, Subject 3

Figure 11: Fourier tract samples of the corpus callosum (*left column*) and fornix (*right column*) of three healthy subjects. For each subject the posterior distributions given the observed datasets were characterised using Metropolis-Hastings Markov-Chain-Monte-Carlo sampling. The semi-transparent outline represents the spatial extent of the Fourier tract samples, while the number of lines approximately represents the Apparent Connection Strength (ACS) (see supplementary material S.5 for a version of this figure with less transparent outlines). Within each of the sample sets, Fourier tracts are assigned unique colours to distinguish them from each other. Images are approximately aligned to the same orientation. Inferred tracts correspond to expected shapes of corpus callosum and fornix.

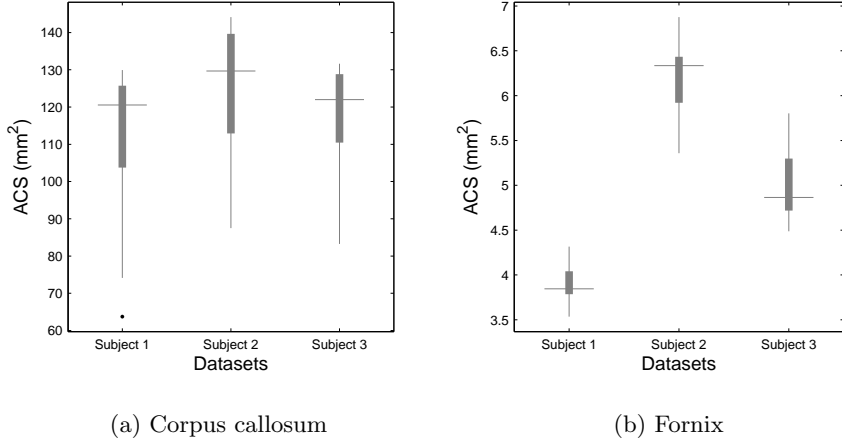


Figure 12: The posterior distribution of Apparent Connection Strength (ACS) of the corpus callosum and fornix estimated from the dMRI datasets of 3 subjects. Given an assumed packing density for the corpus callosum and fornix fibre counts can be estimated from the ACS for the inferred tracts.

The location and orientation of the fibres in the selected phantom neighbourhoods were inferred by FouTS (Figure 9 and 10). However, in some cases the spatial extents of the Fourier tract samples spanned multiple fibre bundles, distorting the WM paths represented by the Fourier tracts. This problem is partially caused by the mismatch in the number of Fourier tracts to reference bundles, i.e. there aren't enough Fourier tracts to represent all the bundles in the phantom section, and partially by the premature convergence of the MCMC chains to local minima. Methods to accurately estimate the optimum number of Fourier tracts, and heuristics to avoid local minima, should be able to improve these distortions but are outside the scope of this study. Nevertheless, despite these issues, FouTS is able to infer the intra-voxel fibre orientations and positioning of fibre bundles within the phantom sections more accurately than existing streamlines-based techniques (Figure 10). This is particularly true near the central voxel of the neighbourhood where the inference is informed by the signal originating from several surrounding voxels. Therefore, FouTS could be used to improve computationally-efficient streamlines methods by more accurate inference of intra-voxel orientations and positions of WM tracts in pre-identified regions of highly complexity.

Applying FouTS to *in vivo* datasets recovers tight peaks in the posterior distribution in regions of the parameter space that correspond to the expected shapes and locations of the seeded tracts. In principle, quantitative bounds on the spatial extent of the WM tracts could be determined to specific levels of confidence from these inferred posterior distributions if

the uncertainties in the signal and noise models were also included within the framework. Such bounds could then be used to quantitatively assess the probability of a WM pathway being resected by a proposed surgical incision, allowing the planned operation to be revised to minimise the associated risks of post-surgical deficits and the remaining risks to be appropriately balanced against the prospective benefits of the procedure.

The Fourier tract model can represent a wide spectrum of realistic WM tract shapes with a low number of parameters. The key assumption of the model is that the constituent pathways of a WM tract are tightly packed and therefore follow similar paths on the macroscopic scale. However, the flexibility of the Fourier tract model still allows the constituent tracts to fan out into a thin ribbon, twist, expand and bunch to reach their targets while circumventing other tracts and brain structures. Some WM structures, such as tracts that divide into sub-branches, cannot be accurately represented by a single Fourier tract, but can be represented by the combination of Fourier tracts (i.e. one for each branch); the automatic detection of these cases is left as future work. Although the inclusion of additional Fourier tracts reduces the efficiency of the representation, the ratio of independent observations to modelled parameters will still be considerably higher than separately representing thin strands of WM. This efficiency of the Fourier tract representation is critical to the success of the proposed Bayesian approach, not only because the reduction in the number of independent fibres compared to thin WM models makes the parameter space faster to search, but also because it reduces the variance in the inferred parameters (Figure 6).

The level of curvature detail of the WM model can be arbitrarily increased by raising the degree (M) of the Fourier tracts. For a given Fourier tract, the optimal choice of M will depend on the SNR of the acquired dMRI dataset and the total signal the Fourier tract represents (i.e. path length \times ACS). However, for typical WM paths, a modest FD degree is sufficient to account for a large proportion of the “power” in the FD parameters (Chung et al., 2010). Furthermore, overspecifying M should have little effect on the inference of the parameters that are supported by the data set because the FD basis functions are orthogonal, i.e. the contribution of a basis function cannot be substituted by a combination of other the basis functions, and the posterior probability density function will be approximately linear at the typically sub-voxel scale at which point further resolution of WM tracts is not possible due to noise. Therefore, in cases where the length the WM tracts are not known *a priori*, a two pass approach can be employed, whereby an initial, “coarse” inference is performed using Fourier tracts of modest degree, which are then substituted by zero-padded versions of themselves with a M that is dependent on their length, before performing another, “fine” inference.

Prior knowledge of the shape of WM tracts is commonly utilised in

streamlines fibre tracking in the form of hard constraints on tract curvature (Johansen-Berg and Behrens, 2009; Jones, 2011). At a minimum, such curvature constraints preclude tracks that fold back on themselves, and they can also aid the resolution of crossing fibres by preventing tracks from taking unrealistically sharp bends to follow the crossing tract. However, the hard curvature constraints must be loose enough to allow tracks to follow bends with high levels of curvature, limiting their utility in resolving complex regions. The proposed framework enables the incorporation of a broader scope of prior information, such as probable densities and smoothness of WM tracts, which could be used to improve the resolution of complex regions of crossing fibres while not unduly biasing the tractography results away from high curvature bends.

The prior term is tuned empirically so that samples drawn from the prior distribution tend to correspond to physically plausible tracts (Figure S.3). Due to the lack of quantifiable data on the curvature and densities of WM tracts with which to properly calibrate the prior term, the overall weight of the prior term is set low relative to the likelihood function, so that the likelihood dominates after the MCMC has converged. In the case of the basic test configurations, the standard deviation of the likelihood function was more than three times that of the prior probability on average. Furthermore, this ratio increases with the size of the image volume so that for the *in vivo* tracts data, the rate of change in the likelihood function from steps in the parameter space is orders of magnitude larger than the corresponding rate of change in the prior probability. Therefore, for WM tracts that span several voxels and above, the prior will have a negligible effect on the shape of the peak in the inferred posterior distribution, even for “hook”-shaped tracts such as the uncinate and fornix. For highly curving WM tracts in smaller regions where the prior may have some influence, a two-pass strategy can be employed where the prior is scaled down or omitted once the MCMC chains have converged to at least part of a tract.

Unlike streamlines tracking and techniques derived from it, where the path at each step only depends on the local dMRI data and on the steps upstream of the current point, the inferred Fourier Tracts take into account both upstream and downstream dMRI data by concurrently fitting the length of the WM pathway (Jbabdi et al., 2007; Zalesky and Fornito, 2009; Kreher et al., 2008). This helps resolve the correct direction of tract curvature, which is ambiguous when only considering the local fibre orientation information, thereby avoiding the dispersion of tracks inherent in streamlines tracking (Jones, 2010). The proposed framework can also be used to infer multiple tracts in the same MCMC chain, thereby including the conditional probability between neighbouring tracts.

Importantly, in addition to providing bounds on the spatial extent of WM tracts, the proposed methods also provide a tangible measure of tract connectivity (ACS), which is easier to interpret than the abstract measures

of connectivity derived from streamlines fibre tracking. The ACS was accurately inferred for the simulated test configurations with the exception of a near constant offset to compensate for the underestimation of the tract lengths due to boundary effects. A critical factor in accurately inferring the tract ACS from *in vivo* data is correctly determining the ratio between the ACS and the diffusion-weighted contrast. While the proposed method for calculating this ratio is sufficiently accurate for the density prior not to interfere with the inference of the tract shape, the intra-subject variation, is on the order of the inter-subject variation suggesting that more sophisticated methods may be required for accurate absolute measurements of ACS.

The accuracy of the signal model is limited by the sparsity of the samples used to approximate the integral over the Fourier tract path in the signal model. For the *in vivo* datasets, the current density of samples prevents the accurate fitting of crossing tracts since the residual signal from the fit of each tract will interfere with the fit of the other. The computation of the signal model is a major bottleneck in the proposed algorithm such that there is a near linear relationship between the number of path samples, T_s and U_s , and the time taken for each MCMC iteration. The MCMC chains used to infer the *in vivo* WM tracts in Figure 11 took on the order of a day. Therefore, significantly increasing the number of samples currently may not be an option for many potential applications. However, since the signal arising from each sample can be calculated independently, multi-threaded or GPU implementations of the signal model calculation will scale near linearly with the number of computational cores, and the resulting time savings could be used to improve the accuracy of the numerical approximations.

Since the dMRI signal is modelled explicitly in the proposed framework, a wide range of acquisition schemes and artefacts can be factored into the Bayesian framework without altering the inference algorithm. For example, the truncated sinc function that is used for the point-spread function can be extended to account for Gibb’s ringing artefacts or convolved with a Lorentzian to model $T2^*$ blurring at the expense of computation time. Since only the anisotropic signal components are used in the likelihood function, any artefacts associated with partial voluming of isotropic tissue such as grey matter and cortico-spinal fluid in the tract inference would be small. Moreover, the only requirement on the configuration of acquired b -values is that the isotropic component of the signal can be accurately estimated, making it possible to use multi-shell HARDI acquisition with the proposed framework (Aganj et al., 2010; Descoteaux et al., 2011).

The proposed formulation of dMRI tractography as a Bayesian problem clearly distinguishes between the physical traits of WM tracts, such as their shape, location and ACS, from the probability of those characteristics occurring, providing results that more directly address many clinical and scientific questions. The high ratio between the number of independent observations and number of parameters in the Fourier tract model leads to

inferences with low variances without the need for strong assumptions on tract curvature. Complete tract data and the conditional probability of surrounding tracts are accounted for in the inference of the WM tracts, thereby aiding the resolution of kissing and crossing configurations, and with more efficient software implementations could be extended to whole brain scale. The proposed methods, therefore not only provide a platform to explore the potential benefits of advanced dMRI acquisition technology for WM tractography, but also represent a robust formulation of the tractography problem based on tangible physical characteristics of WM tracts.

Acknowledgements

Funding for work presented here was supported in part by National ICT Australia (NICTA), Life Sciences division. National ICT Australia is funded by the Australian Government's Department of communications, Information Technology, and the Arts and the Australian Research Council through Backing Australia's Ability and the ICT Research Centre of Excellence programs. We are also grateful to the National Health and Medical Research Council (NHMRC) of Australia, the Australian Research Council (ARC), and the Victorian Government's Operational Infrastructure Support Grant for their support.

Bibliography

- Aganj, I., Lenglet, C., Sapiro, G., Yacoub, E., Ugurbil, K., Harel, N., Aug. 2010. Reconstruction of the orientation distribution function in single- and multiple-shell q-ball imaging within constant solid angle. *Magnetic resonance in medicine* 64 (2), 554–66.
- Alexander, A. L., Hasan, K. M., Lazar, M., Tsuruda, J. S., Parker, D. L., 2001. Analysis of partial volume effects in diffusion-tensor MRI. *Magnetic Resonance in Medicine* 45, 770–780.
- Basser, P. J., Mattiello, J., LeBihan, D., 1994. Estimation of the effective self-diffusion tensor from the NMR spin echo. *Journal of Magnetic Resonance B* 103, 247–254.
- Behrens, T. E., Woolrich, M. W., Jenkinson, M., Johansen-Berg, H., Nunes, R. G., Clare, S., Matthews, P. M., Brady, J. M., Smith, S. M., 2003. Characterization and propagation of uncertainty in diffusion-weighted MR imaging. *Magnetic Resonance in Medicine* 50, 1077–1088.
- Bello, L., Gambini, A., Castellano, A., Carrabba, G., Acerbi, F., Fava, E., Giussani, C., Cadioli, M., Blasi, V., Casarotti, A., Papagno, C., Gupta, A. K., Gaini, S., Scotti, G., Falini, A., Jan. 2008. Motor and language

- DTI Fiber Tracking combined with intraoperative subcortical mapping for surgical removal of gliomas. *NeuroImage* 39 (1), 369–82.
- Bullmore, E., Sporns, O., Feb. 2009. Complex brain networks: graph theoretical analysis of structural and functional systems. *Nature Reviews Neuroscience* 10 (3), 186–198.
- Burkard, R. E., Dell’Amico, M., Martello, S., et al., 2009. *Assignment Problems, Revised Reprint*. Siam.
- Calamante, F., Tournier, J.-D., Jackson, G. D., Connelly, A., 2010. Track-density imaging (TDI): Super-resolution white matter imaging using whole-brain track-density mapping. *NeuroImage* 53 (4), 1233 – 1243.
- Chung, M. K., Adluru, N., Lee, J. E., Lazar, M., Lainhart, J. E., Alexander, A. L., 2010. Cosine series representation of 3d curves and its application to white matter fiber bundles in diffusion tensor imaging. *Statistics and its interface* 3 (1), 69.
- Ciccarelli, O., Catani, M., Johansen-Berg, H., Clark, C., Thompson, A., Aug 2008. Diffusion-based tractography in neurological disorders: concepts, applications, and future developments. *The Lancet Neurology* 7 (8), 715–727.
- Close, T., Tournier, J.-D., Calamante, F., Johnston, L., Mareels, I., Connelly, A., 2009. A software tool to generate simulated white matter structures for the assessment of fibre-tracking algorithms. *NeuroImage* 47 (4), 1288–1300.
- Conturo, T. E., Lori, N. F., Cull, T. S., Akbudak, E., Snyder, A. Z., Shimony, J. S., McKinstry, R. C., Burton, H., Raichle, M. E., 1999. Tracking neuronal fiber pathways in the living human brain. *Proceedings of the National Academy of Sciences of the United States of America* 96, 10422–10427.
- Descoteaux, M., Deriche, R., Le Bihan, D., Mangin, J.-F., Poupon, C., Aug. 2011. Multiple q-shell diffusion propagator imaging. *Medical image analysis* 15 (4), 603–21.
- Farquharson, S., Tournier, J.-D., Calamante, F., Fábinyi, G., Schneider-Kolsky, M., Jackson, G. D., Connelly, A., 2013. White matter fiber tractography: why we need to move beyond DTI. *Journal of neurosurgery* 118 (6), 1367–77.
- Fillard, P., Poupon, C., Mangin, J.-F., 2009. A novel global tractography algorithm based on an adaptive spin glass model. In: *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2009*. Vol. 5761. Springer Berlin / Heidelberg, pp. 927–934.

- Gilks, W., Richardson, S., Spiegelhalter, D., 1996. Markov Chain Monte Carlo in Practice. CRC Press.
- Gong, G., He, Y., Concha, L., Lebel, C., Gross, D. W., Evans, A. C., Beaulieu, C., 2009. Mapping anatomical connectivity patterns of human cerebral cortex using in vivo diffusion tensor imaging tractography. *Cerebral Cortex* 19, 524–536.
- Gudbjartsson, H., Patz, S., 1995. The Rician distribution of noisy MRI data. *Magnetic Resonance in Medicine* 34, 910–914.
- Jbabdi, S., Woolrich, M. W., Andersson, J. L., Behrens, T. E., 2007. A Bayesian framework for global tractography. *NeuroImage* 37, 116–129.
- Johansen-Berg, H., Behrens, T., 2009. Diffusion MRI: From Quantitative Measurement to In-vivo Neuroanatomy. Academic Press.
- Johansen-Berg, H., Behrens, T. E. J., Aug 2006. Just pretty pictures? what diffusion tractography can add in clinical neuroscience. *Current Opinion in Neurology* 19 (4), 379–385.
- Jones, D., 2011. Diffusion MRI: Theory, Methods, and Applications. Oxford University Press.
- Jones, D. K., 2010. Challenges and limitations of quantifying brain connectivity in vivo with diffusion MRI. *Imaging in Medicine* 2 (3), 341–355.
- Kreher, B. W., Mader, I., Kiselev, V. G., Sep 2008. Gibbs tracking: A novel approach for the reconstruction of neuronal pathways. *Magnetic Resonance in Medicine* 60 (4), 953–963.
- Mori, S., Crain, B. J., Chacko, V. P., van Zijl, P. C., 1999. Three-dimensional tracking of axonal projections in the brain by magnetic resonance imaging. *Annals of Neurology* 45, 265–269.
- Neeman, M., Freyer, J. P., Sillerud, L. O., 1991. A simple method for obtaining cross-term-free images for diffusion anisotropy studies in NMR microimaging. *Magnetic Resonance in Medicine* 21, 138–143.
- Nimsky, C., Ganslandt, O., Merhof, D., Sorensen, a. G., Fahlbusch, R., May 2006. Intraoperative visualization of the pyramidal tract by diffusion-tensor-imaging-based fiber tracking. *NeuroImage* 30 (4), 1219–29.
- Parker, G. J., Haroon, H. A., Wheeler-Kingshott, C. A., 2003. A framework for a streamline-based probabilistic index of connectivity (PICO) using a structural interpretation of MRI diffusion measurements. *Journal of Magnetic Resonance Imaging* 18, 242–254.

- Reisert, M., Mader, I., Anastasopoulos, C., Weigel, M., Schnell, S., Kiselev, V., 2011. Global fiber reconstruction becomes practical. *NeuroImage* 54 (2), 955 – 962.
- Schreiber, J., Riffert, T., Anwander, A., Knösche, T. R., 2014. Plausibility Tracking: A method to evaluate anatomical connectivity and microstructural properties along fiber pathways. *NeuroImage* 90, 163–78.
- Sherbondy, A., Dougherty, R., Ananthanarayanan, R., Modha, D., Wandell, B., 2009. Think global, act local; projectome estimation with bluematter. In: *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2009*. pp. 861–868.
- Smith, R. E., Tournier, J.-D., Calamante, F., Connelly, A., 2013a. Evidence for the improved biological interpretability of white matter connectivity derived following tractogram filtering using SIFT. In: *Proceedings of the International Society of Magnetic Resonance in Medicine*. p. 2135.
- Smith, R. E., Tournier, J.-D., Calamante, F., Connelly, A., 2013b. SIFT : Spherical-deconvolution informed filtering of tractograms. *NeuroImage* 67, 298–312.
- Tournier, J.-D., Calamante, F., Connelly, A., 2007. Robust determination of the fibre orientation distribution in diffusion MRI: non-negativity constrained super-resolved spherical deconvolution. *NeuroImage* 35, 1459–1472.
- Tournier, J.-D., Calamante, F., Connelly, A., Mar. 2012. MRtrix: Diffusion tractography in crossing fiber regions. *International Journal of Imaging Systems and Technology* 22 (1), 53–66.
- Tournier, J.-D., Calamante, F., Gadian, D. G., Connelly, A., 2004. Direct estimation of the fiber orientation density function from diffusion-weighted MRI data using spherical deconvolution. *NeuroImage* 23, 1176–1185.
- Tournier, J.-D., Yeh, C. H., Calamante, F., Cho, K. H., Connelly, A., Lin, C. P., 2008. Resolving crossing fibres using constrained spherical deconvolution: Validation using diffusion-weighted imaging phantom data. *NeuroImage* 42, 617–625.
- Tuch, D. S., 2004. Q-ball imaging. *Magnetic Resonance in Medicine* 52, 1358–1372.
- Wedeen, V. J., Hagmann, P., Tseng, W. Y., Reese, T. G., Weisskoff, R. M., 2005. Mapping complex tissue architecture with diffusion spectrum magnetic resonance imaging. *Magnetic Resonance in Medicine* 54, 1377–1386.

Zalesky, A., Fornito, A., 2009. A DTI-derived measure of cortico-cortical connectivity. *IEEE Transactions on Medical Imaging* 28 (7), 1023 –1036.

Appendix

A. All empirically tuned parameters of the Bayesian framework

Symbol	Description	Value
$\kappa_f^{(0)}$	Primary frequency prior scalar	15
$\kappa_f^{(1)}$	Auxiliary frequency prior scalar	60
κ_h	Hook prior scalar	10^5
κ_D	High density prior scalar	1
κ_d	Low density prior scalar	0.01
κ_b	Barrier prior scalar	10^6
q_b	Barrier prior exponent	10
ϵ_l	ACS lower bound from primary magnitude	0.0025
ϵ_w	ACS lower bound from auxiliary magnitude	0.0025
T_h	Hook prior number of samples along primary axis	100
U_h	Hook prior number of samples along auxiliary axis	15
T_d	Density prior number of samples along primary axis	100
T_s	Signal model number of samples along primary axis	25
U_s	Signal model number of samples along auxiliary axis	4

B. Implementation of prior probability density function

B.1. Frequency prior

The frequency prior is designed to make Fourier tracts with large high frequency oscillations improbable in the proposed framework (Chung et al., 2010). The expression for the frequency prior is straightforward to implement in the Fourier tract model as the frequency of the oscillations increase with the FD order. However, the frequency of an oscillation in a FD path also depends on the parameterisation speed of the path. Therefore, the frequency prior is formulated in terms of the ratio between the magnitude of higher order vectors and the the linear component magnitude. A Gaussian distribution is assumed for this ratio for each higher order vector ($m > 1$),

$$\log [P_f(\Theta)] = - \sum_{m=2}^M m \left(\kappa_f^{(0)} \frac{\|\boldsymbol{\theta}_m^{(0)}\|^2}{\|\boldsymbol{\theta}_1^{(0)}\|^2} + \kappa_f^{(1)} \sum_{l=1}^2 \frac{\|\boldsymbol{\theta}_m^{(l)}\|^2}{\|\boldsymbol{\theta}_1^{(0)}\|^2} \right), \quad (17)$$

where the standard deviations of the distributions for the primary and auxiliary parameter vectors are scaled by empirically tuned constants $\frac{1}{\kappa_f^{(0)}}$ and $\frac{1}{\kappa_f^{(1)}}$ respectively (Appendix A), and decay with $\frac{1}{m}$ to match the decay in the average reconstruction error with increasing degree for FDs fit to typical DTI streamlines in Chung et al. (2010).

B.2. Hook prior

Figure 4b shows another type of tract configuration that is assumed to be improbable, where a WM path “hooks” back on itself to reach its endpoint. A FD is considered to form a “hook” whenever a point on its path projects onto the line joining its endpoints past either endpoint, in which case one of the angles at the base of the triangle formed between such a point on the FD path and the two endpoints, ϕ_0 and ϕ_1 , exceeds 90° (Figure 13). The hook prior is designed so that the probability of a Fourier tract configuration decreases sharply after ϕ_0 or ϕ_1 exceed 90° at any point across its volume, sampled at T_h positions of τ along the tract and on a $U_h \times U_h$ grid of μ and ν ,

$$\log [P_h(\Theta)] = -\frac{\kappa_h}{T_h Z_h} \sum_{v=-U_h}^{U_h} \sum_{u=-U_h}^{U_h} \sum_{t=1}^{T_h} \sum_{e=0}^1 (1 - \cos[\phi_{e,t,u,v}(\Theta)])^2 \quad (18)$$

where

$$\begin{aligned} \cos [\phi_{e,t,u,v}(\Theta)] &= \frac{\|\mathbf{w}_{e,t,u,v}(\Theta) \cdot \mathbf{w}_{b,t,u,v}(\Theta)\|}{\|\mathbf{w}_{e,t,u,v}(\Theta)\| \|\mathbf{w}_{b,t,u,v}(\Theta)\|}, \\ e &= 0, 1, \quad t = 1, \dots, T_h, \quad u = -U_h, \dots, U_h, \\ v &= -U_h, \dots, U_h, \quad \frac{u^2 + v^2}{U_h^2} < 1, \end{aligned} \quad (19)$$

and

$$\mathbf{w}_{0,t,u,v}(\Theta) = \gamma \left(\frac{t}{T_h + 1}, \frac{u}{U_h}, \frac{v}{U_h}, \Theta \right) - \gamma \left(0, \frac{u}{U_h}, \frac{v}{U_h}, \Theta \right), \quad (20)$$

$$\mathbf{w}_{1,t,u,v}(\Theta) = \gamma \left(1, \frac{u}{U_h}, \frac{v}{U_h}, \Theta \right) - \gamma \left(\frac{t}{T_h + 1}, \frac{u}{U_h}, \frac{v}{U_h}, \Theta \right), \quad (21)$$

$$\mathbf{w}_{b,t,u,v}(\Theta) = \gamma \left(1, \frac{u}{U_h}, \frac{v}{U_h}, \Theta \right) - \gamma \left(0, \frac{u}{U_h}, \frac{v}{U_h}, \Theta \right). \quad (22)$$

κ_h is empirically tuned to ensure that unrealistic hooks are improbable while not unduly biasing the overall distribution away from configurations with realistic levels of curvature near their endpoints (Appendix A), and Z_h is the number of samples across the auxiliary axes such that $\frac{u^2 + v^2}{U_h^2} < 1$.

B.3. Density prior

The density priors represent the expectation regarding both the upper bounds on the density of the constituent fibres within WM tracts due to physical constraints, and the lower bounds on that density for what could reasonably be called a single tract. The density of a Fourier tract at a position τ along the Fourier tract is the ratio of its ACS with the cross-sectional area, $A(\tau, \Theta)$. The cross-sectional area is determined by the projection of

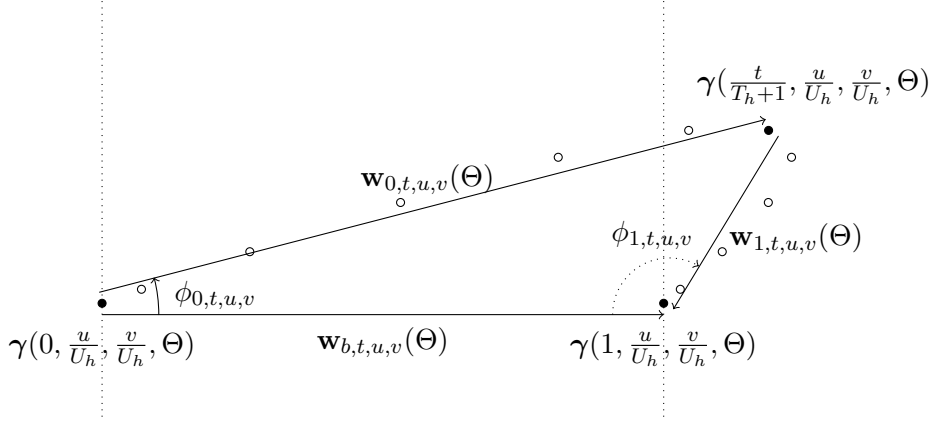


Figure 13: The angles of the triangle formed between the endpoints of a Fourier descriptor (FD) and a sample along the FD’s path that are used to determine whether the FD path forms a “hook”.

the cross-product between the auxiliary axes onto the orientation of the central path,

$$A(\tau, \Theta) = \frac{\pi}{4} \frac{\gamma'(\tau, 0, 0, \Theta)}{\|\gamma'(\tau, 0, 0, \Theta)\|} \cdot \gamma(\tau, 1, 0, \Theta) \times \gamma(\tau, 0, 1, \Theta). \quad (23)$$

By definition, WM tracts will be typically close to unit density, with decreases in density expected in sections where tracts interdigitate with other tracts or spread out to contact grey matter targets. The expression for the density prior is

$$\log [P_d(\Theta, \alpha)] = -\frac{1}{T_d} \sum_{t=1}^{T_d} \kappa_D (1 - D_t(\Theta, \alpha))^2 + \kappa_d \left(1 - \frac{1}{D_t(\Theta, \alpha)}\right)^2 \quad (24)$$

where

$$D_t(\Theta, \alpha) = \frac{\eta(\Theta, \alpha)}{A\left(\frac{t}{T_d+1}, \Theta\right)}, \quad (25)$$

and κ_D and κ_d tuned empirically so that Fourier tracts with densities significantly greater than unity are improbable while significantly lower densities are still probable (Appendix A).

B.4. Barrier prior

The barrier prior places a soft constraint on the location of the Fourier tracts within the Bayesian framework by assuming an extremely low probability of sections of the Fourier tracts lying outside of the region of interest. Like the hook prior, the log of the barrier prior is the sum of contributions from samples throughout the Fourier tract volume. The contribution of each sample to the barrier prior is its displacement from the image centre

normalised by the barrier length, which is then raised and scaled by the empirically chosen exponent $2q_b$ and scalar κ_b respectively (Appendix A),

$$\log [\mathbf{P}_b(\Theta)] = -\frac{\kappa_b}{T_h Z_h} \sum_{v=-U_h}^{U_h} \sum_{u=-U_h}^{U_h} \sum_{t=1}^{T_h} \sum_{d=1}^3 \left(\frac{\gamma_d(\tau, \mu, \nu, \Theta) - c_d}{R_d} \right)^{2q_b} \quad (26)$$

where c_d and R_d are the centre of the image and the displacement from the centre to the barrier respectively along the d th dimension of the image. q_b and κ_b are tuned such that the \mathbf{P}_b has little effect on the posterior distribution when the Fourier tract section lies within the barrier but a large effect when the section lies outside the barrier.

C. Point-spread function

For images acquired on a Cartesian grid the full point-spread function is the product of the point spread functions along each axis. The signal contribution of a path segment is then this point-spread function multiplied by ACS of the tract and the estimated SIAR, Ξ , which is estimated for each dataset,

$$H(\mathbf{p}) = \Xi \prod_{q=1}^3 h_q(p_q). \quad (27)$$

For 1D and 2D dMRI acquisitions, where $d = 3$ is the slice direction, the ideal in-plane point-spread function is a normalised sinc function due to the Fourier reconstruction of the acquired signal,

$$h_q(p_q) = \text{sinc} \left(\frac{p_q - x_q}{\Delta x_q} \right), \quad q = 1, 2. \quad (28)$$

where x_d and Δx_d are the voxel location and dimension along the q^{th} axis respectively. The point-spread function along the slice selection axis when using an apodized selection pulse is approximated by the truncated polynomial function,

$$h_d(p_3) = \begin{cases} \left(\frac{p_3 - x_3}{\Delta x_3} \right)^4 - 2 \left(\frac{p_3 - x_3}{\Delta x_3} \right)^2 + 1, & |p_3 - x_3| < \Delta x_3 \\ 0, & \text{otherwise} \end{cases} \quad (29)$$

Supplementary Material

S.1. Gaussian approximation of Rician noise

The effect of substituting a Gaussian distribution noise model for the expected Rician distribution noise distribution (Gudbjartsson and Patz, 1995), for a typical 60 diffusion-encoding ($b=3000 \text{ s/mm}^2$) scheme was explored over SNR values 2, 5, 10, 20. A set of diffusion-weighted signals was simulated for the encoding scheme from a single segment of a WM path. For each

SNR value, 1000 weighting sets were generated from the base set by applying different Rician noise realisations. The likelihood functions for both the Rician and Gaussian noise models were sampled over the 180° rotation of a matching path segment for each noise realisation, assuming a matching SNR. The likelihood function was then normalised to give the posterior probability density function (assuming a uniform prior). 99% confidence bounds were determined for each noise realisation and the difference between the bounds for Gaussian and Rician noise models recorded. For each SNR value the average confidence bound for both noise models plus the average difference between them was calculated.

Figure S.1 shows the normalised posterior distributions for the different SNR levels. For the SNR=10 and SNR=20 cases, there is minimal difference between the Gaussian and Rician noise models particularly in the high probability regions of the curve (Figure S.1c & S.1d). For the SNR=5 case, there is a significant underestimation of the probability when substituting a Gaussian model for a Rician model in the low probability regions of the curve but there little difference in the high probability regions, reflected in the negligible, 0.2° difference in the 99% confidence bounds between the models (Figure S.1b). The average difference in the confidence intervals when the SNR=2 (Figure S.1a) is 22° . However for SNR=2, the average confidence interval for the Rician model was 65.9° and therefore of little utility.

S.2. Assessment of crossing fibre angle resolution

The angular resolution of the likelihood function for a typical 60 direction, $b = 3000 \text{ s/mm}^2$ encoding scheme was examined via the independent rotations of two infinitesimally short segments of modelled WM. Four “observed” images were each simulated using the forward signal model arising from two tract segments, with angles of incidence between the segments of 15° , 22.5° , 30° and 45° in the four images respectively. In each case the first segment was aligned at 0° . While no noise was applied to the simulated “observed” image, an SNR of 15 was assumed for the likelihood function. Samples of the likelihood function are taken over two parameter dimensions, the independent in-plane rotations of two segments, on a 100×100 grid from -90° to 90° relative to the first segment of the observed image.

The raster plots of the likelihood function for the orientation of two path segments with varying angular separations show the angular resolution limits of a typical 60 direction $b=3000 \text{ s/mm}^2$ diffusion encoding scheme (Figure S.2). For the 45° crossing case, the likelihood function has two distinct, isotropic peaks centred about the two “true” configurations, noting that the true configurations are identical under exchange of the path segments (Figure S.2a). As the crossing angle decreases to 30° the peaks start to noticeably elongate towards each other (Figure S.2b). The elongation becomes more pronounced in the 22.5° crossing case, where the two peaks merge at a contour within an order of magnitude of their maximum (Figure S.2c).

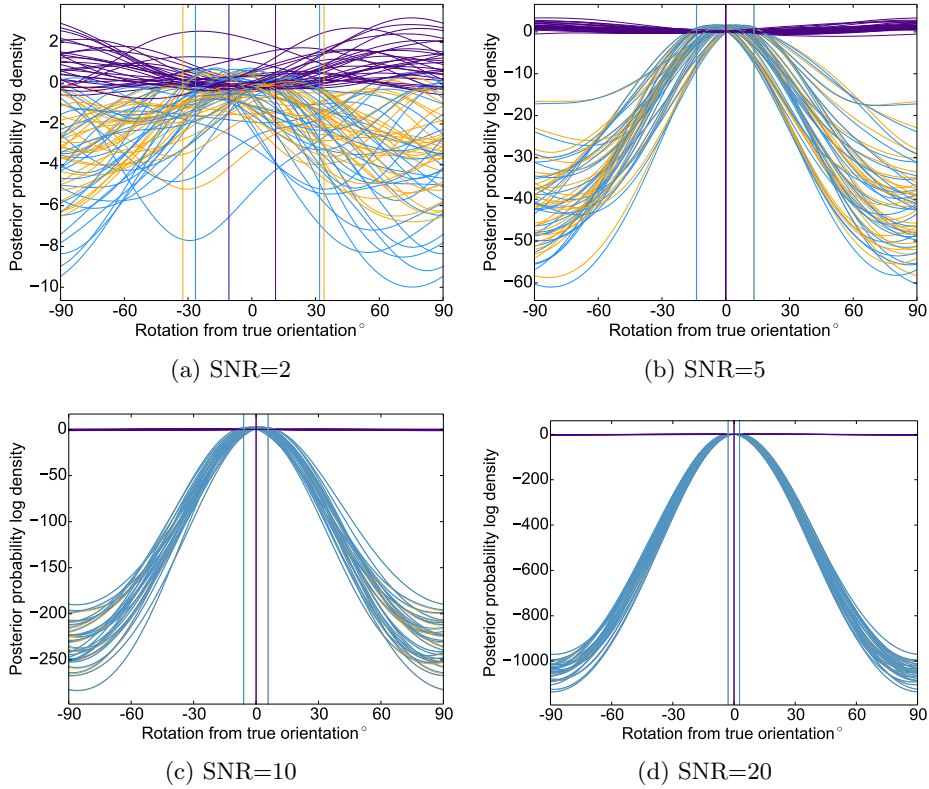


Figure S.1: Posterior probability log density function of angular deviation between true and sample segments for different noise levels. (a) SNR 2. (b) SNR 5. (c) SNR 10. (d) SNR 20. For each noise realisation the log-likelihood is calculated with an assumed Rician noise model (orange) and a Gaussian approximation (blue), with the pairwise difference between the Rician and Gaussian curves (i.e. for same noise realisation) shown in purple. In each case the assumed SNR corresponds to the level of noise added to the simulated reference signal. The 99% confidence intervals averaged over all noise realisations are shown by the vertical orange and blue lines for the Rician and Gaussian models respectively and the average pairwise difference between these intervals. For $\text{SNR} \geq 5$ the difference between the Gaussian and Rician models is negligible within the high probability section of the curve, which is of interest.

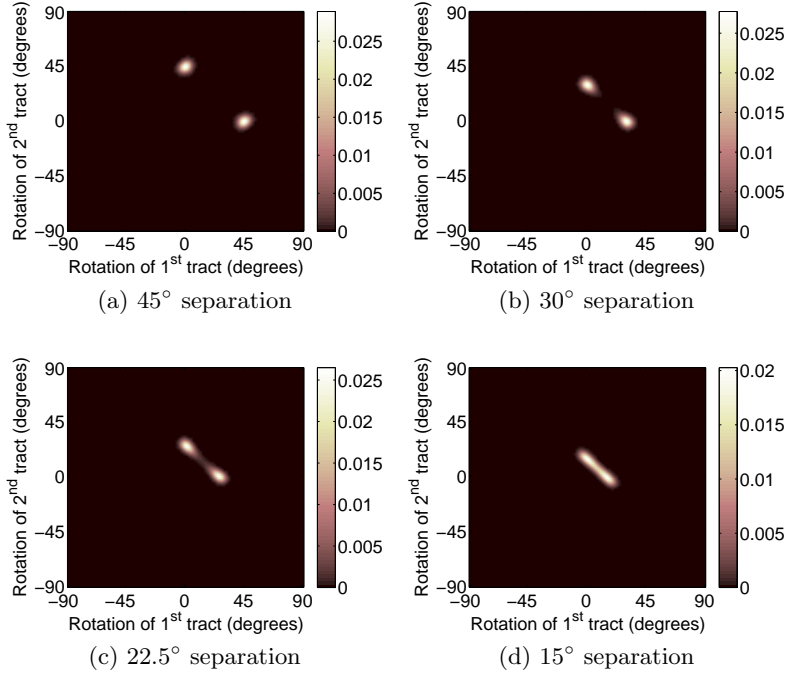


Figure S.2: Peaks in the likelihood function around the estimation of two distinct fibre orientations at differing levels of angular separation. Peaks in the likelihood function correspond to the “true” configuration used to generate the dataset and the “mirror” configuration where the indices of the segments are exchanged (a) Peaks around true and mirrored configurations in the 45° case are clearly separated. (b) The peaks around the true and mirrored configurations begin to extend towards each other as the angle of separation is lowered to 30° . (c) The peaks around the true and mirrored configurations in the 22.5° case begin to merge. (d) The peaks around the true and mirrored configurations in 15° case have merged into a single elongated peak.

When the crossing angle is 15° , the two peaks are no longer distinguishable from each other and merge to form one long peak, such that the midway point between the true configurations—where the two paths run parallel at 7.5° —is almost as likely as the true 15° crossing (Figure S.2d).

S.3. Assessment of a priori tract shapes

To confirm that probable regions in the prior distribution correspond to common-sense physical characteristics of WM tracts, the posterior distributions of single Fourier tracts within a single encoding direction datasets simulated with white noise were sampled. In this case, since there were no salient features in the dataset, the morphologies of the Fourier tract samples were largely determined by the shape of the prior distribution.

The final sample of nine independent MH-MCMC chains of 100,000 iterations were selected from a posterior distribution over a single Fourier tract

($M=6$). In place of an observed data set, a $10 \times 10 \times 10$ image with 1 encoding direction was simulated by applying Gaussian noise with a standard deviation 0.04 times $b = 0$ signal. Since the widths of the Fourier tracts were not constrained by the image data in this case, an additional Gaussian prior distribution was placed on the Fourier tract ACS for a consistent scale. The parameters of $\theta_1^{(0)}$, $\theta_0^{(1)}$ and $\theta_0^{(2)}$, were initialised from normal distributions of $\mathcal{N}(0, 0.1)$, $\mathcal{N}(0, 0.025)$ and $\mathcal{N}(0, 0.025)$ respectively (units are relative to isotropic voxel size), and all other parameters were initialised to 0.

The samples generated from the pure Gaussian noise images correspond to a broad spectrum of physically plausible tracts demonstrating the effect of the prior distribution (Figure S.3). Although each Fourier tract is defined by 64 parameters ($M=6$), each sample has a complex combination of features such as bending, fanning and twisting. None of the Fourier tracts have high curvature or density, and therefore they correspond well to the expected shapes of WM tracts. While some of the Fourier tract samples bend significantly at several points along their length, the overall path of the tracts are smooth and they do not exhibit high frequency oscillations.

S.4. Approximations to the sinc point-spread function

The effect on the likelihood function of approximating the EPI point-spread function in the 2D imaging plane was studied for four cases: the ideal (untruncated) sinc function, a sinc function truncated at the width of 1 voxel from its peak in all directions,

$$f_s = \begin{cases} \frac{\sin[x]}{x}, & x < 1 \\ 0, & \text{otherwise} \end{cases} \quad (30)$$

a quartic polynomial function,

$$f_p = \begin{cases} x^4 - 2x^2 + 1, & x < 1 \\ 0, & \text{otherwise} \end{cases} \quad (31)$$

where x is the distance from the centre of the voxel normalised by the voxel length, and a top-hat function

$$f_t = \begin{cases} 1, & x < 1/2 \\ 0, & \text{otherwise} \end{cases} \quad (32)$$

Note that the top-hat function is the equivalent of the typical voxel-wise fibre estimations (Tuch, 2004; Tournier et al., 2008; Wedeen et al., 2005).

An observed $9 \times 9 \times 3$ image volume is simulated for a straight FD that is aligned with the z axis and passes through the volume. A simulated observed image is generated using the signal model and parameters are the same as those used in rest of the paper with the exception of the point-spread function is extended to cover the 9×9 plane. Samples of the likelihood

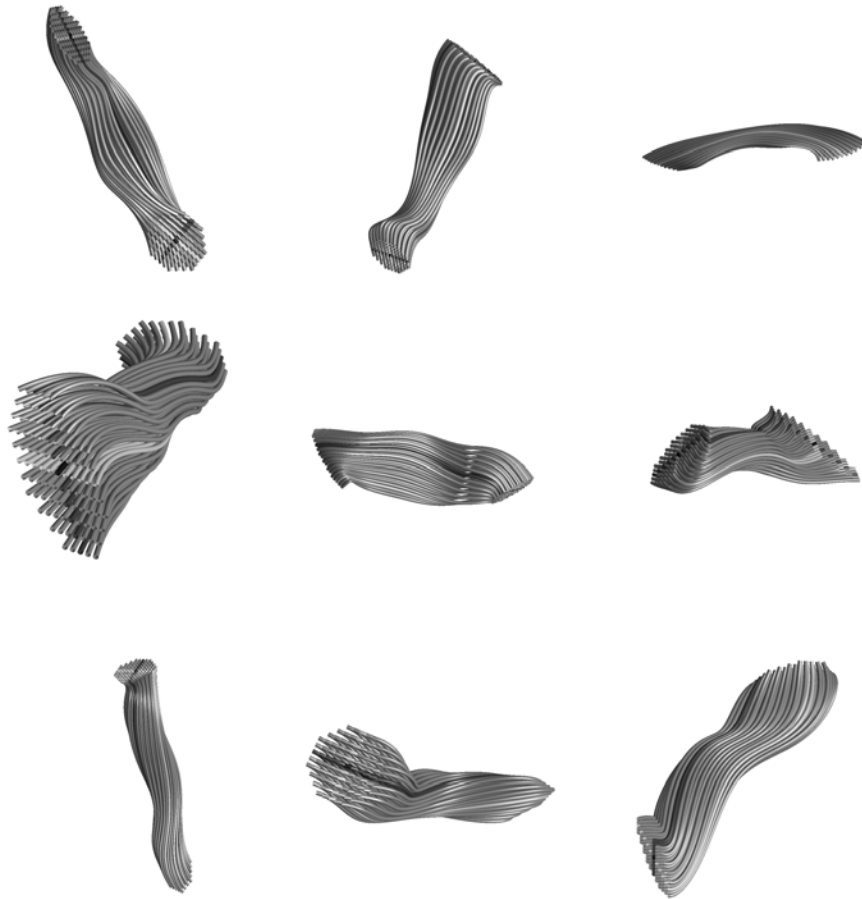


Figure S.3: Final states of 9 MCMC chains generated from the prior distribution. The primary FD is marked in black and the first and second auxiliary axes are marked light grey and darker grey respectively. The Fourier tract samples exhibit complex combinations of bending, fanning and twisting. The Fourier tract samples are plausible tract shapes as they do not have sections of high curvature, high density or high frequency oscillations.

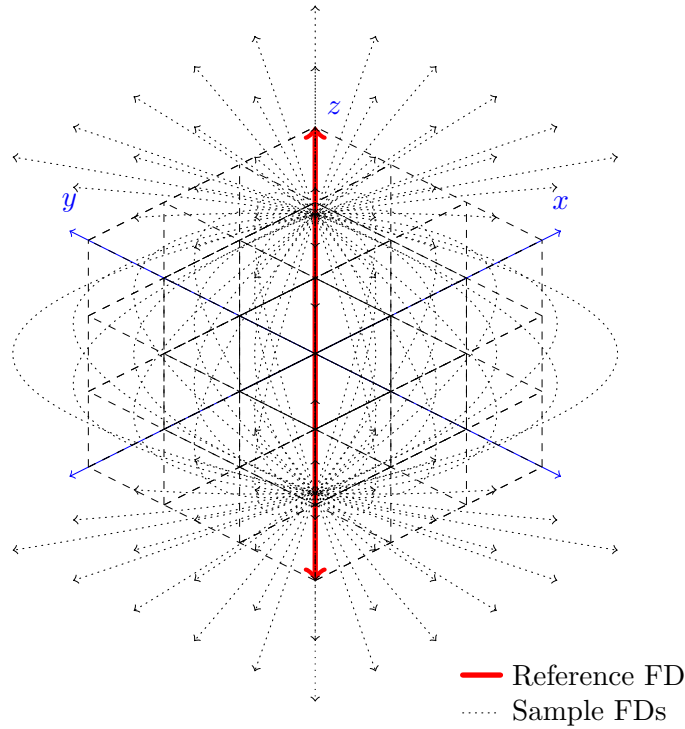


Figure S.4: Sweep of $m = 2$ parameters of a FD. FD parameter $\theta_{2,1}$ is varied along the first dimension of the sweep causing the strand to curve the x axis, and $\theta_{2,2}$ is varied along the second dimension of the sweep resulting in the strand curving along the y axis.

function are taken for FD samples with the same $m = 0, 1$ parameters as the reference strand but varying $m = 2$ parameters perpendicular to the axis of the reference strand (Figure S.4). The likelihood is sampled over a regular grid for the $m = 2$ parameters for each of the four point-spread function cases (Figure S.5).

With the exception of the top-hat point-spread function case (Figure S.5d), where the combination of many strand fragments causes a tessellation in the log-likelihood map, the relationship between the θ_2 curvature parameters and the likelihood is smooth and without local maxima. The log-likelihood of the untruncated sinc point-spread function forms a near quadratic bowl (Figure S.5a), whereas in the truncated sinc (Figure S.5b) and the polynomial (Figure S.5c) cases the peak is slightly stretched along the diagonals. The smooth likelihood function in these cases highlights the benefits of the direct relationship between the strand parameters and fragment positions/orientations in the proposed model; the strand parameters are related to the fragment positions and tangents via a linear transform.

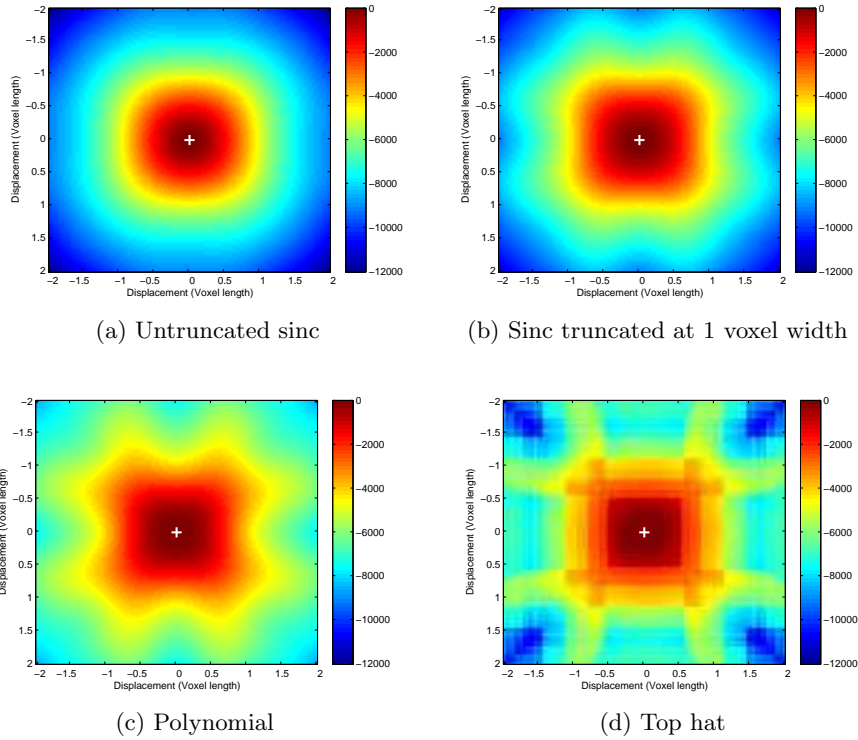
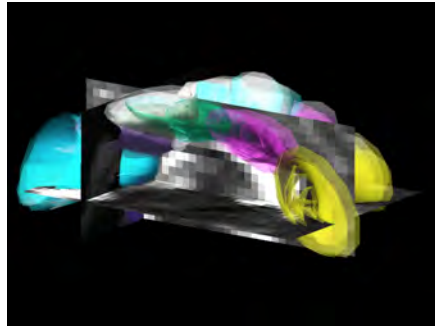


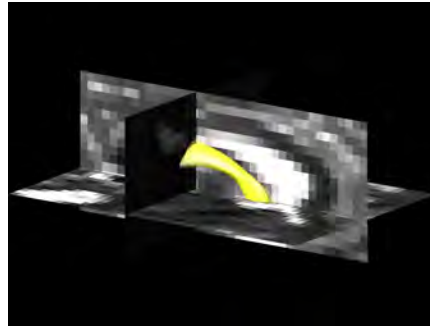
Figure S.5: Map of the log-likelihood as a function of the $m = 2$ parameters of a FD. A $9 \times 9 \times 3$ image region is simulated for a straight reference FD, centred at the origin along the z axis. Samples are taken by varying $\theta_{2,1}$ and $\theta_{2,2}$ so the FD curves along the x and y axis respectively. Four cases are shown corresponding to differently modelled point-spread functions for the images simulated for the samples. (a): the peak in untruncated sinc case forms a near quadratic bowl. (b & c): the peaks in the truncated sinc and the polynomial cases are distorted towards the corners of the map. (d): the peak in top-hat case is tessellated due to the effect of the discontinuity in the top-hat function in conjunction with the sampled fragments used to numerically integrate along the strand.

S.5. Reduced transparency of in vivo sampling figures

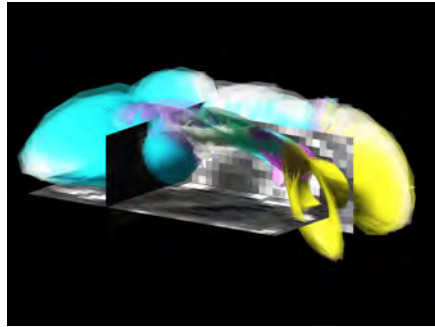
Figure 11 is reproduced in Figure S.6 without the internal lines representing the ACS and with reduced transparency of the tract convex hull to highlight the spatial extent of the inferred tract samples.



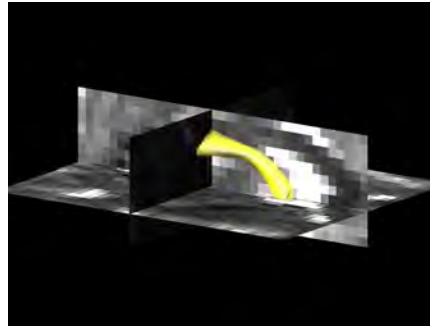
(a) Corpus callosum, Subject 1



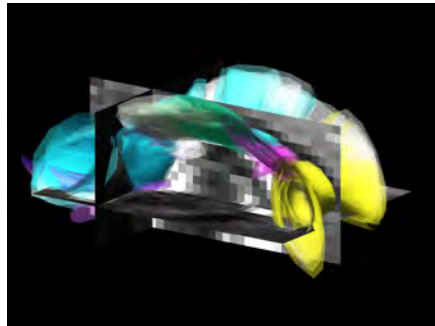
(b) Fornix, Subject 1



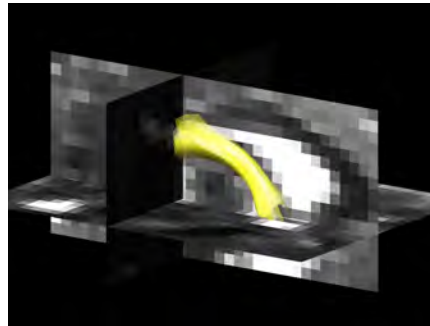
(c) Corpus callosum, Subject 2



(d) Fornix, Subject 2



(e) Corpus callosum, Subject 3



(f) Fornix, Subject 3

Figure S.6: Fourier tract samples of the corpus callosum (*left column*) and fornix (*right column*) of three healthy subjects. For each subject the posterior distributions given the observed dMRI datasets were characterised using Metropolis-Hastings Markov-Chain-Monte-Carlo sampling. The semi-transparent outline represents the spatial extent of the Fourier tract samples. Images are approximately aligned to the same orientation. Within each of the sample sets, Fourier tracts are assigned unique colours to distinguish them from each other.