

# CATERPillar: A Flexible Framework for Generating White Matter Numerical Substrates with incorporated Glial Cells

Jasmine Nguyen-Duc<sup>a,b,1</sup>, Malte Brammerloh<sup>a</sup>, Melina Cherchali<sup>a</sup>, Inès de Riedmatten<sup>a,b</sup>, Jean-Baptiste Péro<sup>a</sup>, Jonathan Rafael-Patiño<sup>c</sup>, and Ileana O Jelescu<sup>a,b</sup>

<sup>a</sup>Department of Radiology, Lausanne University Hospital (CHUV), Lausanne, Switzerland

<sup>b</sup>Faculty of Biology and Medicine, University of Lausanne (UNIL), Lausanne, Switzerland

<sup>c</sup>Signal Processing Laboratory 5 (LTS5), Ecole Polytechnique Federale de Lausanne (EPFL), Lausanne, Switzerland

<sup>1</sup>Corresponding author: jasmine.nguyen-duc@chuv.ch

June 20, 2025

## Abstract

Monte Carlo diffusion simulations in numerical substrates are valuable for exploring the sensitivity and specificity of the diffusion MRI (dMRI) signal to realistic cell microstructure features. A crucial component of such simulations is the use of numerical phantoms that accurately represent the target tissue, which is in this case, cerebral white matter (WM). This study introduces CATERPillar (Computational Axonal Threading Engine for Realistic Proliferation), a novel method that simulates the mechanic of axonal growth using overlapping spheres as elementary units. CATERPillar facilitates parallel axon development while preventing collisions, offering user control over key structural parameters such as cellular density, tortuosity, beading and myelination. Its uniqueness lies in its ability to generate not only realistic axonal structures but also realistic glial cells, enhancing the biological fidelity of simulations. We showed that our grown substrates feature distributions of key morphological parameters that agree with those from histological studies. The structural realism of the astrocytic components was quantitatively validated using Sholl analysis. Furthermore, the time-dependent diffusion in the extra- and intra-axonal compartments accurately reflected expected characteristics of short-range disorder, as predicted by theoretical models. CATERPillar is open source and can be used to (a) develop new acquisition schemes that sensitise the MRI signal to unique tissue microstructure features, (b) test the accuracy of a broad range of analytical models, and (c) build a set of substrates to train machine learning models on.

## 1 Highlights

1. CATERPillar generates realistic axons and glial cells that may co-exist within numerical substrates.
2. Synthetic axons had similar morphologies to those segmented from human electron microscopy in previous works.
3. The morphological features of synthetic astrocytes closely matched those observed in rodent histology.
4. The functional form of diffusion time-dependence in the intra- and extra-axonal spaces agreed with experimentally observed disorder power laws in voxels composed of synthetic axons.

## 2 Keywords

Diffusion; MRI; Time dependence; White matter; Microstructure; Monte Carlo simulation; Cell structure

### 3 Introduction

Brain white matter (WM) consists of a dense network of axons originating from neurons whose cell bodies are mostly located in gray matter (GM). Many of these axons are encased in concentric layers of proteolipid membranes known as myelin, which serve to insulate them and enhance the speed and efficiency of electrochemical signal conduction [1]. Axons serve as the communication gateway of the brain, allowing WM to function as the “edges” within the brain’s structural network [2]. Other cell types, such as glial cells, are also present in WM. These glial cells do not directly contribute to neuronal signalling, but perform essential support functions: oligodendrocytes promote myelination, astrocytes maintain the optimal chemical environment for neuronal communication and contribute to neurovascular coupling [3], and microglia clear cellular debris from injury sites [4, 5]. Both axons and glial cells have highly dynamic morphologies and are influenced by their local environment [6]. To investigate the microstructure of these cells in WM, classical histology has proven to be fundamental to modern neuroscience as it reveals mechanisms of neural connectivity and communication, and exposes the cellular origins of neurodegenerative disease [7]. However, these methods are expensive, invasive, and destructive to the tissue, restricting their use to small samples from post-mortem specimens [7]. Nonetheless, thanks to their high level of anatomical detail, classical histological methods remain a reliable gold standard for tissue microstructure.

Diffusion magnetic resonance imaging (dMRI) has garnered interest within the scientific community as it shows potential for investigating brain tissue microstructure non-invasively, as opposed to classical histological methods. As the diffusion signals are sensitive to the random displacement of water molecules, we are able to study tissue on scales lower than image resolution (tens of micrometers vs millimeters) [8]. The ability of dMRI to non-invasively capture features of intact tissue, combined with a broad field of view and relatively fast, cost-effective acquisition, makes it particularly well-suited for in vivo brain studies [7]. The growing interest for accurate estimations of microstructural features from dMRI signals has driven the development of a wide range of analytical models [7, 9–11]. These models rely on simplifying assumptions to describe relevant geometric and diffusion-related characteristics. Nevertheless, the accuracy of these models and the robustness of their estimates to microstructural complexity and to deviations from the assumptions cannot be easily verified due to the absence of realistic ground truth (GT) microstructural parameters. New and robust validation approaches are thus essential.

One such approach is numerical simulations in synthetic substrates, which offer a flexible and cost-effective way to study the dMRI signal in realistic yet controlled environments. They elucidate a link between GT microstructural features and dMRI signals, in complex environments for which analytical solutions cannot be obtained [12]. Briefly, to model water molecule trajectories and produce synthetic dMRI signals, a substrate representing the tissue of interest is first generated, after which Monte Carlo (MC) diffusion simulation tools [13–17] are employed. This framework opens several possibilities: (i) to test and optimise the sensitivity of various diffusion acquisition schemes to tissue microstructure features for which no analytical formulation exists (e.g. branching, beading, and spines), (ii) to test the accuracy and robustness of a broad range of analytical models in substrate conditions that deviate from their assumptions, and (iii) to build a set of substrates to train machine learning models on, alleviating the need for analytical models altogether. Notably, numerical simulations have been previously utilised to explore DWI changes in response to dynamic physiological processes such as ischemic stroke [18] and neuronal activation [19].

Various approaches for generating WM numerical substrates exist. These approaches either derive meshes directly from segmented histological data [20–24] or comprise generative models that produce substrates “from scratch”, aiming to match GT tissue parameters derived from histology [5, 13, 25–28]. In the first case, the substrates are undeniably realistic, while the available configurations are constrained by specific histological samples. Amongst the frameworks that replicate WM tissue “from scratch” are MEDUSA [26], CONFIG [27], CACTUS [28], and WMG [5]. Each of these methods prioritises different aspects, such as computational efficiency, packing density and realism of the growth process. However, while some of these frameworks support the generation of glial cells (some more elaborate than others), their biological realism has never been evaluated. Such validation efforts have been exclusively focused on synthetic axons.

Here, we introduce a WM substrate generator entitled CATERPillar (Computational Axonal Threading Engine for Realistic Proliferation), which combines various strengths of previous generators. We go beyond the state-of-the-art by placing a strong emphasis not only on generating anatomically realistic axons and their myelin sheath, but also on incorporating biologically accurate glial cells. CATERPillar integrates the use of overlapping unit spheres from MEDUSA [26], which are computationally efficient,

and the natural fibre growth approach from CONFIG [27]. As parallelised axonal growth is enabled, the run-time for substrate generation is considerably decreased compared to most previous generators. Additionally, CATERPillar provides precise control over axonal tortuosity and beading amplitude, while allowing to vary the number of fibre populations. A user-friendly GUI will be publicly released to make this novel tool accessible to all users, regardless of their programming experience. Also, we intend to release the code open-source so that users may modify it and/or suggest modifications which could improve this tool. The aim of this work is to validate the realism of our generated axons by comparing their features with EM-derived distributions. Additionally, the diffusion properties within each compartment were tested against theoretical predictions from effective medium theory [29]. As for glial cells, astrocytes were generated using morphological parameters derived from previous histological studies in rodents, and several features were compared against those reported in the literature.

## 4 Methods

### 4.1 Growing WM substrates with CATERPillar

In the following, we first outline the main steps of the substrate growth algorithm and describe its implementation. Next, we provide a more detailed description of the creation of the different substrate constituents: glial cells, axons and myelin.

#### 4.1.1 Substrate growth implementation

The substrate is generated in a box-shaped volume through a multi-step growth process implemented in C++ programming language, following user-defined parameters governing the substrate growth (Figure 1). The substrate generation is performed as follows :

1. Glial cell somas are placed within the volume until a target volume fraction is achieved.
2. Axons are grown from one plane towards the opposite plane, with their trajectories shaped by parameters such as desired packing density, tortuosity, beading, and fibre orientation distribution function (fODF).
3. Glial processes extend from the somas, growing while avoiding collisions with existing axons, until the desired volume fraction is reached.

All cells are constructed by progressively placing overlapping spheres in chains. The spacing between two adjacent spheres is initially set to  $\max(R_1, R_2)$  for efficiency, where  $R_1$  is the radius of the first sphere and  $R_2$  the radius for the next. To smooth the surface, additional overlapping spheres are added between consecutive ones, spaced by  $d = \frac{\max(R_1, R_2)}{F}$ , where  $F$  is a user-defined overlap factor. Higher  $F$  values improve realism but increase computational cost. Supplementary materials (section 14.1) detail how  $F = 4$  balances accuracy and computation time.

**Development of the GUI:** To improve the ease of use of the CATERPillar substrate generator, a GUI was developed in C++ using the Qt framework. It provides users with an intuitive and interactive environment to adjust the parameters that govern substrate generation. Additionally, the GUI supports the visualisation of the generated substrates, enabling users to inspect their spatial structure and composition. Figure 2 illustrates an example of a generated substrate containing both axons and astrocytes, showcasing the tool’s capability to render complex multicellular environments.

#### 4.1.2 Axons

**Initial growth process:** The numerical substrate is a cube of edge length  $L$ . In order to generate axons in this cube, a set of radii are drawn from a representative Gamma distribution (user-defined by parameters  $\alpha$  and  $\beta$ ) [30, 31], until the ratio of their total disk areas to the voxel  $x$ - $y$  plane area ( $L^2$ ) achieves the target density  $f$ . Each drawn radius is attributed to an axon, with a starting position placed randomly on a plane and an “attractor” on the opposite one. At each step of the growth process, successive spheres are sequentially positioned in a chain following the attractor’s direction. The axon grows until it reaches the opposite plane or collides with another obstacle. If an obstacle is encountered, multiple attempts are made to identify an alternative available position for the next sphere. If a suitable

A)	Parameters	Control names	Value ranges	Default values
	Gamma distribution parameters for inner axons	$\alpha$	$[0, \infty]$	4
		$\beta$	$[0, \infty]$	0.12
	Volume fraction [%]	f	$[0, 100]$	30
	Beading amplitude	A	$[0, 1]$	0.3
	Tortuosity	$\varepsilon$	$[0, \infty]$	0.4
	Orientation dispersion	$c_2$	$[0.75, 1]$	1
	Myelin sheath	$C_0, C_1, C_2$	$[0, \infty]$	0.35, 0.006, 0.024
	Number of populations	N	$[1, 3]$	1

B)	Parameters	Control names	Value ranges	Default values
	Mean soma radius [ $\mu\text{m}$ ]	$R_{\text{soma}}$	$[0, \infty]$	5
	Standard deviation for soma radius [ $\mu\text{m}$ ]	$\sigma_{\text{soma}}$	$[0, \infty]$	0.5
	Mean process length [ $\mu\text{m}$ ]	$L_{\text{process}}$	$[0, \infty]$	30
	Standard deviation for process length [ $\mu\text{m}$ ]	$\sigma_{\text{process}}$	$[0, \infty]$	15
	Number of primary processes	$N_{\text{process}}$	$[1, \infty]$	10
	Soma volume fraction [%]	$f_{\text{soma}}$	$[0, 100]$	1
	Process volume fraction [%]	$f_{\text{process}}$	$[0, 100]$	6
	Branching allowed ?	N/A	yes/no	yes

Figure 1: List of parameters used to generate populations of axons (A) and glial cells (B) in CATERPillar.

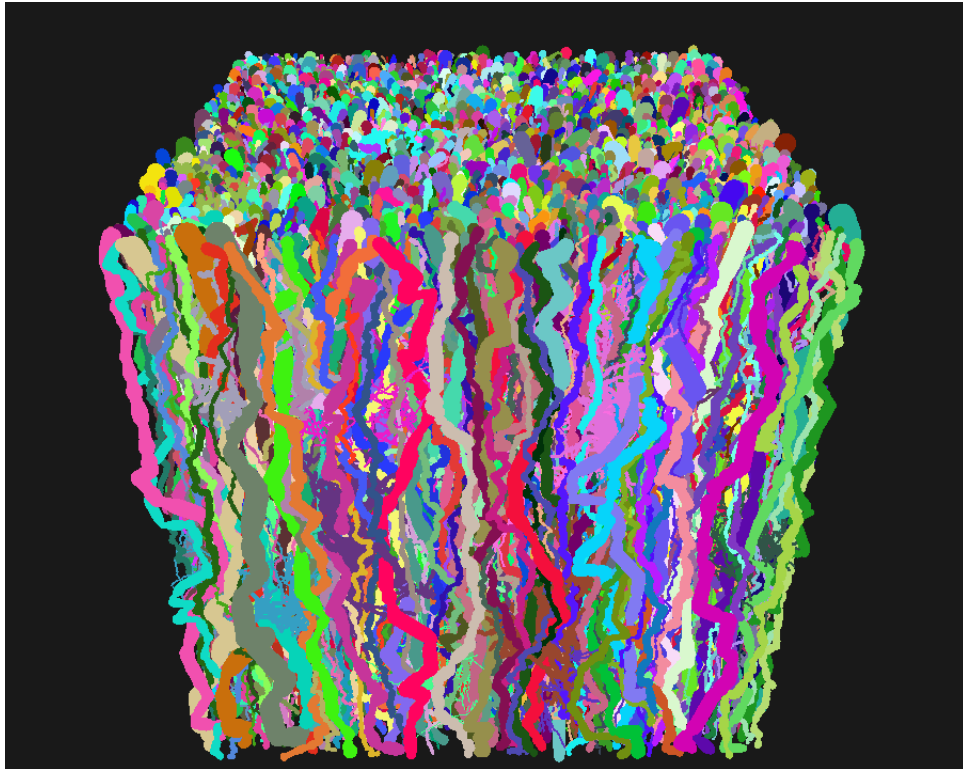


Figure 2: Example of a numerical substrate shown with the GUI's visualisation tool.

position cannot be found after several attempts, the radius of the sphere is gradually reduced to enable the axon to pass through tighter regions. This reduction typically affects only a small sequence of consecutive spheres and is therefore not expected to significantly alter the axon’s average radius. However, if growth remains obstructed despite these adjustments, the axon is deleted, and axon growth is reinitiated at a different starting position. The collision detection algorithm used is inspired by the Sweep and Prune method [32]. This algorithm enhances efficiency compared to the brute-force approach by creating bounding boxes around each object. To boost the speed of the algorithm even further, we use threads to grow axons in parallel.

**Control over tortuosity:** During growth, the next sphere’s position is determined using polar and azimuthal angles, both sampled from Gaussian distributions centred at 0 to bias growth along the z-axis, before rotating toward the attractor. The standard deviation  $\epsilon$ , which is user-defined, controls deviation from the attractor’s direction, and thus affects the axon tortuosity. The latter is defined as the total axon length divided by the straight-line distance between endpoints. Prior studies report tortuosities around 1–1.2 for various axons [24, 33]. We recommend  $\epsilon = 0.4$  to reproduce a tortuosity of 1.2. Figure 3A shows how  $\epsilon$  influences tortuosity.

**Multiple fibre populations:** Multiple fibre populations can be grown, with primary bundle directions perpendicular to one another. For two populations ( $N = 2$ ), the user can specify whether the fibres grow as adjacent, parallel sheets or intersect perpendicularly in an interwoven pattern. For three populations ( $N = 3$ ), the fibres are exclusively interwoven (see Figure 3B).

**fODF:** Users can specify the fibre orientation distribution function by setting the value  $c_2$ , which is the mean squared cosine of axon orientations relative to the bundle direction, based on the Watson distribution. The  $c_2$  parameter is converted to the concentration parameter  $\kappa$ , which defines the initial orientation of each axon relative to its attractor. A  $c_2 = 1/3$  yields an isotropic configuration, while  $c_2 = 1$  yields perfect alignment. For a single fibre population, the minimum  $c_2$  that can be achieved is 0.75. To obtain lower values, more fibre populations are necessary. Figure 3C illustrates substrates with varying  $c_2$ .

**Radius modulation:** Axon radius variation can be introduced by sampling each new sphere’s radius from a Gaussian distribution centred on the previous one. The latter is constrained within  $r \pm A \cdot \bar{r}$ , where  $r$  is the first sphere’s radius,  $\bar{r}$  the initial axon radius from the Gamma distribution, and  $A$  controls the amplitude of variation. This keeps average axon calibre consistent while allowing local fluctuations.

#### 4.1.3 Myelin sheath

Once axonal growth is complete, myelin can be added by introducing spheres within the existing ones. This is achieved by duplicating each sphere and reducing its radius. The region between the inner and outer spheres constitutes the myelin compartment, whose thickness is determined by a user-specified g-ratio. This g-ratio ( $\frac{R_{in}}{R_{out}}$ ) is calculated as follows, assuming that myelin thickness follows a log-linear relationship as previously established by Berthold et al. (1983) [34]:

$$M = C_0 + C_1 \cdot (2R_{in}) + C_2 \cdot \ln(2R_{in}) \quad (1)$$

where  $M$  is the myelin thickness and  $R_{in}$  is the inner radius. A previous study using EM data [35] proposed a fitted model with parameters  $C_0 = 0.35$ ,  $C_1 = 0.006$ , and  $C_2 = 0.024 \mu m$ . In CATERPillar, users may specify the parameters  $C_0$ ,  $C_1$ , and  $C_2$  (see Figure 1); however, we recommend keeping the default values to ensure that the generated myelin sheaths remain consistent with biologically observed structures. A lower bound of the g-ratio is imposed, ensuring that it does not fall below 0.2 for any sphere, in line with [35]. The user may define and grow a population of myelinated axons and a population of unmyelinated axons, as the two may coexist within the same voxel.

#### 4.1.4 Glial cells

First, glial somas are placed in the volume using radii sampled from a user-defined Gaussian distribution (mean =  $R_{soma}$  and standard deviation =  $\sigma_{soma}$ ). Once the target for the glial soma volume fraction ( $f_{soma}$ ) is reached, axons grow while avoiding these somas, as previously described. Each soma then

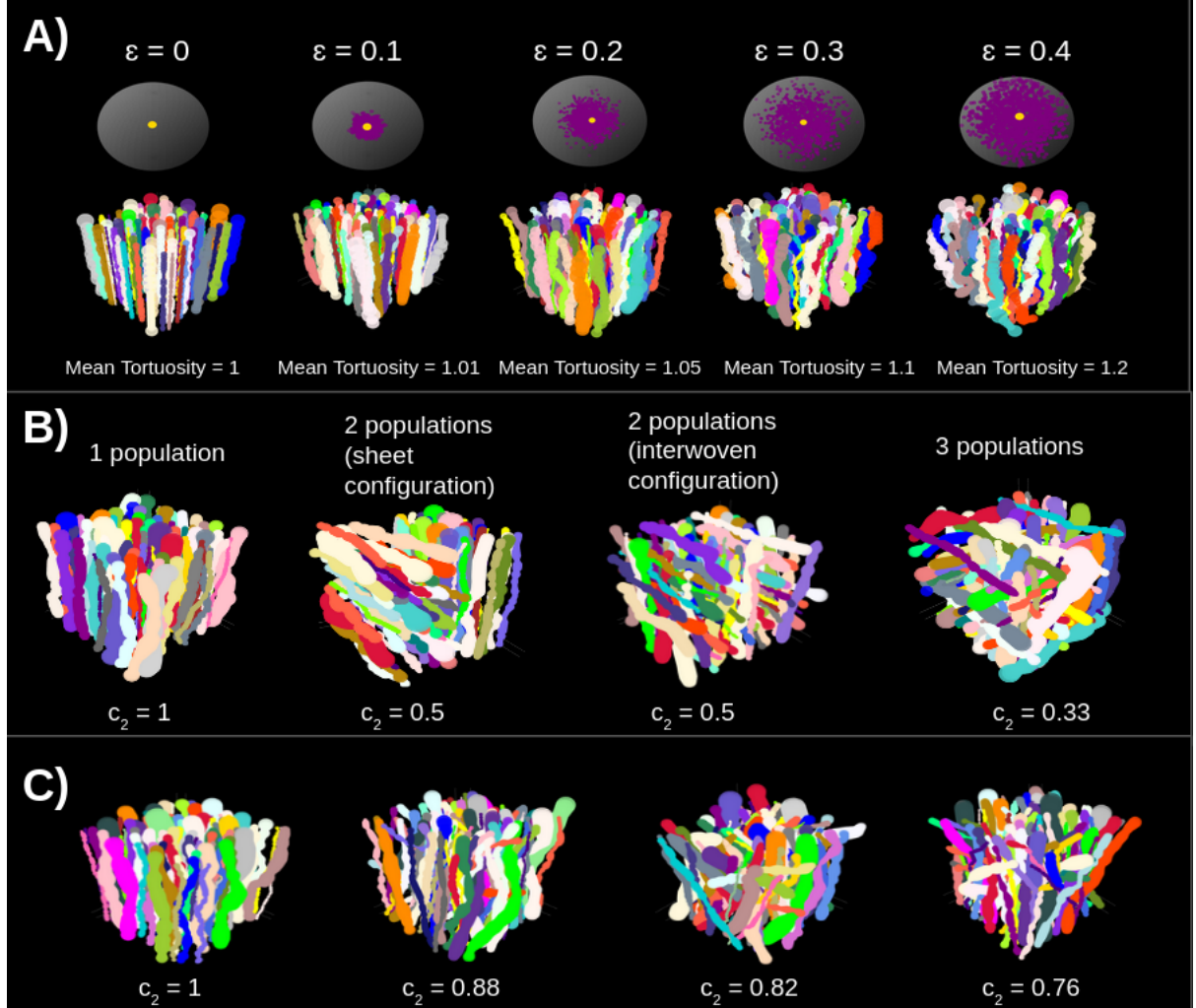


Figure 3: Illustrations of numerical substrates (width =  $30 \mu m$ ) with variations in specific parameters. A) Increasing the  $\epsilon$  parameter leads to an increase in tortuosity values. B) Different number of fibre populations that cross each other perpendicularly. In case of 2 populations, they cross each other in two different ways: sheet configuration or interwoven configuration. C) The angles between the starting points of axons and their attractors can be set so that a specific orientation dispersion is obtained, leading to the desired  $c_2$ .

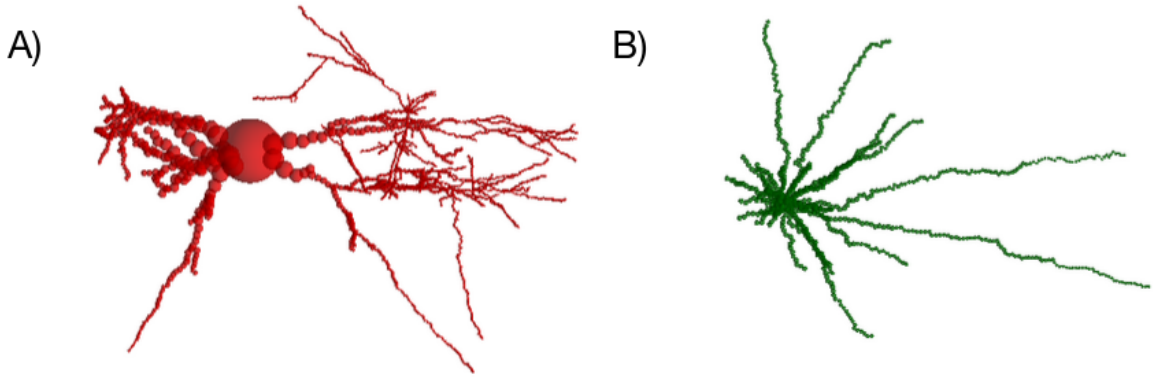


Figure 4: Example of a synthetic glial cell with primary and secondary processes (A) and a synthetic glial cell with only primary processes. (B).

extends a user-defined number of primary processes ( $N_{process}$ ), defined as the branches originating directly from the soma. If branching is enabled, higher-order processes may sprout from existing ones until a target volume for processes ( $f_{process}$ ) is reached; otherwise, additional primary processes grow (see Figure 4). Processes grow in random directions, with their radii decaying exponentially with length. A decrease in the diameter of processes with their length was observed in microscopic images of glial cells in [36]. We model this decrease as an exponential decay, as this was shown to be a good fit [37]. Growth continues until either a specific process length is reached or when an obstacle cannot be avoided. The length for each process is drawn from a Gaussian distribution with mean  $L_{process}$  and standard deviation  $\sigma_{process}$ . It is defined here as the cumulative distance from the soma to the terminal tip, encompassing not only primary processes but also all higher-order branches. By measuring the full path from the soma rather than just individual branch segments, we ensure that higher-order processes are represented with proportionally shorter lengths compared to the primary processes.

Our tool allows for generating up to two distinct glial cell populations within a single voxel. Each population can be independently configured with its own set of parameters. This enables the inclusion of various glial populations (such as astrocytes and oligodendrocytes) or the representation of healthy and pathological glial cells within the same substrate.

## 4.2 Morphological analysis

We analysed the morphology of both axons and glial cells in our numerical substrates. To evaluate the realism of myelinated axons, we generated a representative WM substrate with 1176 axons grown in a  $(100\ \mu\text{m})^3$  voxel using default parameters shown in Figure 1A. Inner radii followed a Gamma distribution ( $\alpha = 4, \beta = 0.12$ ), and g-ratios were locally derived using a fitted model from [35]. The beading amplitude was set to 0.3 and  $\epsilon$  to 0.4. Resulting distributions for inner/outer diameters, g-ratio and diameter coefficients of variation (CV) were compared to previous findings on axons segmented from electron-microscopy, specifically to Figure 4 from [35].

For glial cells, substrates of  $(100\ \mu\text{m})^3$  were generated based on morphological parameters representative of rodent astrocytes, as detailed data for these cells are widely available [38–47]. The parameters were set to the default values shown in Figure 1B. The analysis was performed on astrocytes grown in two settings: one within axon-filled substrates (79 astrocytes) with a volume fraction  $f$  of 30%, and another in empty space (64 astrocytes). As our synthetic astrocytes are expected to be adaptive to their environment (in this case, to the presence of axons), we expect to find morphological differences between the two groups, despite them having grown with the same set of parameter values. For clarity, we refer to these as protoplasmic and fibrous astrocytes, as these are the names given to astrocytes present within GM and WM tissue respectively.

To demonstrate these differences, we applied Sholl analysis to quantify the arbour complexity of both astrocyte types by counting the number of intersections between astrocytic processes and concentric spheres centred on the soma. This Sholl analysis gave insight on the critical radius, process maximum



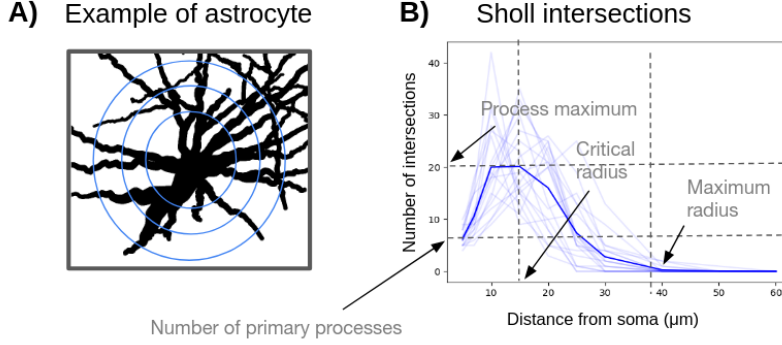


Figure 5: A) Illustration showing a synthetic astrocyte with concentric circles around its soma, depicting the concept behind Sholl analysis. B) In Sholl analysis, the process maximum refers to the highest number of intersections between astrocytic processes and the concentric circles. The critical value denotes the radial distance from the soma at which this maximum occurs, while the maximum radius corresponds to the furthest extent of the astrocytic processes [48]. Individual Sholl intersection curves for each glial cell in a voxel (light lines) are averaged to produce the darker mean curve.

and maximum radius, as illustrated in Figure 5. For each astrocyte, we also determined the total number of processes, the mean process radius, as well as the mean and total process length. This was done when taking into account only the primary processes, and for all processes combined. The statistical analysis between protoplasmic and fibrous astrocytes was conducted using the unpaired two-sample t-test in Python, and significance was assumed when  $p < 0.05$ . Additionally, we compared the fractional anisotropy (FA) that was computed as described in [47]. Each process contributed to the FA of its corresponding glial cell, with its influence weighted according to its volume. We expect fibrous astrocytes to exhibit greater elongation compared to those in empty space, resulting in higher FA values. By confirming this in our analysis, we also confirm that our cells are adaptive to their environment through their collision avoiding rules during growth.

### 4.3 Monte Carlo simulations of water diffusion

After generating WM substrates and confirming their morphological accuracy, we evaluated how accurately they reflected dMRI signals. To do so, we conducted MC simulations within the substrates containing impermeable and non-myelinated axons and examined whether the time dependence of diffusivity aligned with theoretical predictions from effective medium theory [29]. Before doing so, we adapted a previously published MC simulator to the specifics of the CATERPillar substrate generation.

#### 4.3.1 Adaptations and Extensions of the MC/DC tool

The diffusion simulator used in this work was the MC/DC (Monte Carlo Diffusion and Collision) simulator, a previously published open-source tool [15]. MC/DC integrates core dMRI protocols with a range of tissue models and supports various substrate geometries, including cylindrical, spherical, and mesh-based structures. Thanks to its open-source nature, MC/DC has been successfully adapted in multiple prior studies [28, 49]. For the CEXI model [49], the tool was extended to incorporate membrane permeability in permeable spherical cells.

For this study, we extended MC/DC to implement mirror boundary conditions. MC/DC originally relied on periodic tiling, where structures repeat seamlessly across boundaries. While effective for many scenarios, this approach is less suitable for substrates with high orientation dispersion. We introduced mirror boundary conditions, where random walkers cross voxel boundaries and diffuse within the mirrored substrate on the other side. Simulating diffusion within a mirrored environment has the advantage of maintaining microstructural continuity [12]. However, it can increase the apparent fODF and introduce curvature artifacts. To minimise such edge effects, we placed walkers within a central region buffered from the edges by a distance of  $30 \mu\text{m}$ . This distance should be at least equal to the mean squared displacement:  $\langle x^2 \rangle = 2Dt$ , where  $D$  is diffusivity and  $t$  is diffusion time. Additionally, to support intracellular diffusion, we modified MC/DC to allow walkers to diffuse across overlapping spheres by ignoring collisions at surfaces that lie within neighbouring spheres.



### 4.3.2 Time dependence of diffusivity

To verify that the CATERPillar-generated substrates display the short-range disorder observed in histological data, Diffusion Kurtosis Imaging (DKI) was computed using the TMI toolbox [50–52] across multiple diffusion times, separately for the intra-axonal space (IAS) and extra-axonal space (EAS). Theoretically, if the substrate exhibits short-range disorder as expected, the IAS diffusivity along the axons should scale as  $1/\sqrt{t}$ , where  $t$  is the diffusion time, while the EAS diffusivity perpendicular to the axons is expected to follow a  $\ln(t)/t$  dependence. For further details on the short-range disorder, refer to the Supplementary Materials.

Simulations were performed on substrates of  $150\ \mu\text{m}$  per side, a size selected based on previous findings showing that the standard deviation of the diffusion signal plateaus at voxel sizes around  $100\ \mu\text{m}$  [15, 53]. The IAS volume fraction was approximately 50%, and the axons were unmyelinated yet impermeable, in a single bundle, without glial cells. Three types of substrates were generated: (i) substrate with axons modelled as cylinders (ii) one with straight and beaded axons (beading amplitude = 0.3), (iii) one with tortuous ( $\epsilon = 0.4$ ) and beaded axons. The radii of axons were drawn from a Gamma distribution with  $\alpha = 4$  and  $\beta = 0.25$ .

Simulations used bulk diffusivities of  $2.5\ \mu\text{m}^2/\text{ms}$  for IAS and  $1.5\ \mu\text{m}^2/\text{ms}$  for EAS. Random walkers were initialised in a central  $(90\ \mu\text{m})^3$  region within the  $(150\ \mu\text{m})^3$  cube to avoid edge effects. A step size of  $0.1\ \mu\text{m}$  yielded 115,500 and 69,300 steps for IAS and EAS, respectively. Simulations used  $10^5$  walkers per compartment, repeated three times per substrate to capture variability.

The analysis focused on diffusion times of 11, 16, 25, 44, 109 and 157 ms as similarly done in [23]. To approximate a narrow-pulse Pulse Gradient Spin Echo (PGSE) sequence, the gradient duration  $\delta$  was fixed at 4 ms, while the inter-gradient duration  $\Delta$  and gradient strength  $G$  varied to obtain different combinations of diffusion times and b-values. Sixty-one unique gradient directions per shell were generated with optimal angular coverage using the method outlined by [54]. The b-values utilised were 0, 500, 1000 and  $2000\ \text{s}/\text{mm}^2$ . Diffusion tensors were estimated from the simulated IAS and EAS signals separately, from which AD within the IAS and RD within the EAS were derived and analysed as a function of diffusion time.

## 5 Results

### 5.1 Performance of WM substrate generation with CATERPillar

Figure 6 shows the time necessary to grow a substrate of  $150\ \mu\text{m}$  in length for various axonal packings, when setting  $\epsilon$  to 0.4 and the beading amplitude to 0.3. By parallelising the axonal growth computations across 20 threads in this example, high packing densities could be achieved for this voxel size under 12 hours. Each thread used one core and approximately 300 MB of RAM.

### 5.2 Morphological analysis

#### 5.2.1 Myelinated Axons

We evaluated our synthetic axon morphologies against histological measurements illustrated in Figure 4 from [35], and the results were highly consistent. Figure 7 shows histograms of inner and outer diameters, g-ratios, diameter CV, and tortuosity for both our synthetic axons. The near-perfect alignment of these distributions with those in Figure 4 from [35] indicated that our generation pipeline faithfully reproduced real axonal geometry. Specifically, our g-ratios tracked the expected inner-diameter dependence from Equation 1, staying between 0.2 and 1, and the CV peaked around 0.2 for outer diameters and 0.3 for inner diameters. Tortuosity, which was not reported in [35], followed an almost Gaussian profile centred at 1.18 (range 1.05–1.4), matching physiological values reported by [33]. These close correspondences confirmed that our CATERPillar-generated substrates effectively captured the realistic structure of WM axons.

#### 5.2.2 Astrocytes

Figure 8A and B show the two types of astrocyte-containing substrates: one with astrocytes grown in empty space and the other with astrocytes grown in a voxel populated with axons. The averaged Sholl intersection curves in panel C, computed across all astrocytes within each voxel, reflected their characteristic arbour complexity. The curve for both synthetic protoplasmic and fibrous astrocytes

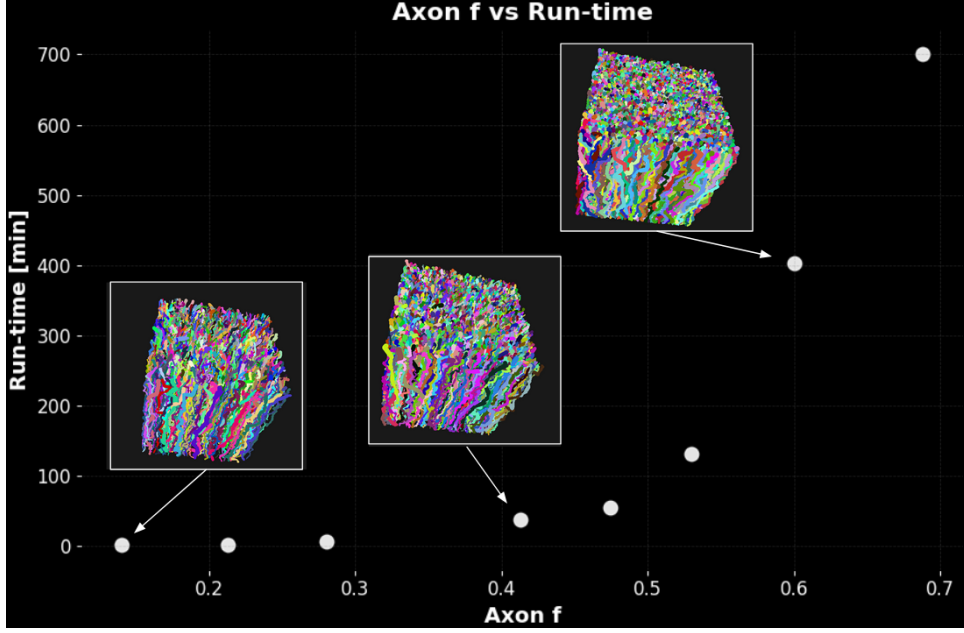


Figure 6: The run-time versus the axonal packing  $f$  for substrates of  $150 \mu m$  in length.

closely matched previous histological findings in rodent GM [38, 41–44, 46, 55], showing the typical increase in intersections up to a process maximum between 20 and 50 intersections (y-axis), at a critical radius between 20 and  $30 \mu m$  (x-axis), followed by a gradual decline ending between 40 and  $60 \mu m$ . Fibrous astrocytes exhibited a higher maximum process length than protoplasmic astrocytes, while their critical radius was slightly smaller.

In Figure 9, further morphological analysis examined the number of processes, total process length, mean process length, and mean process radius for both astrocyte types. These measurements were assessed separately for primary processes (panels A-D) and for all processes (panels E-H). These results agreed well with previous histological findings, illustrated with a pink boxes [41, 42, 44, 46, 47, 55]. In panel A, both astrocyte types exhibited a median close to 10 primary processes, consistent with the input specification. In panel C, the median primary process length for both astrocyte types fell below the  $30 \mu m$  input target, reflecting possible premature terminations of growth when collisions with obstacles occurred. Although we did not explicitly control the features shown in panels B, C, F, and G, they nonetheless remained within the expected ranges derived from rodent histological protoplasmic data. Panel D demonstrates that the mean radius of primary processes exceeded that of all processes (panel H), consistent with the fact that process radius decreases as length increases. Overall, the morphological differences indicated that fibrous astrocytes possessed a greater number of processes that were, on average, shorter than those of protoplasmic astrocytes.

### 5.3 Time dependence of diffusivity

To assess whether the structural disorder of the generated substrates fell within the short range disorder, the time dependence of AD in the IAS and RD in the EAS were analysed across the three different substrate configurations, each composed of impermeable and non-myelinated axons (Figure 10).

For both substrates containing beaded axons, the time-dependence of the AD in IAS was better described by the function  $1/\sqrt{t}$  ( $R^2 = 0.92 - 0.93$  in Figure 10A) than by  $1/t$  ( $R^2 = 0.84 - 0.88$  in Figure 10B), indicating that media with beaded axons are likely to be within the short-range disorder class as expected. In contrast, for the straight cylindrical axons, AD remained constant, indicating no time dependence. Assuming short-range disorder,  $D_\infty$  can be approximated from the intercepts of the linear fits in Figure 10A for each substrate. The highest value was observed for straight axons ( $2.53 \mu m^2/ms$ ), followed by straight beaded axons ( $2.14 \mu m^2/ms$ ), and tortuous beaded axons ( $1.39 \mu m^2/ms$ ). The slopes of these fits, referred to as restriction strengths in [23], were 0.05 for cylinders, 0.44 for straight beaded axons, and 0.41 for tortuous beaded axons. These results suggest greater diffusion restriction in beaded axons compared to cylinders, though somewhat unexpectedly, restriction appears slightly stronger in straight beaded axons than in their tortuous counterparts.

## Morphology of synthetic axons

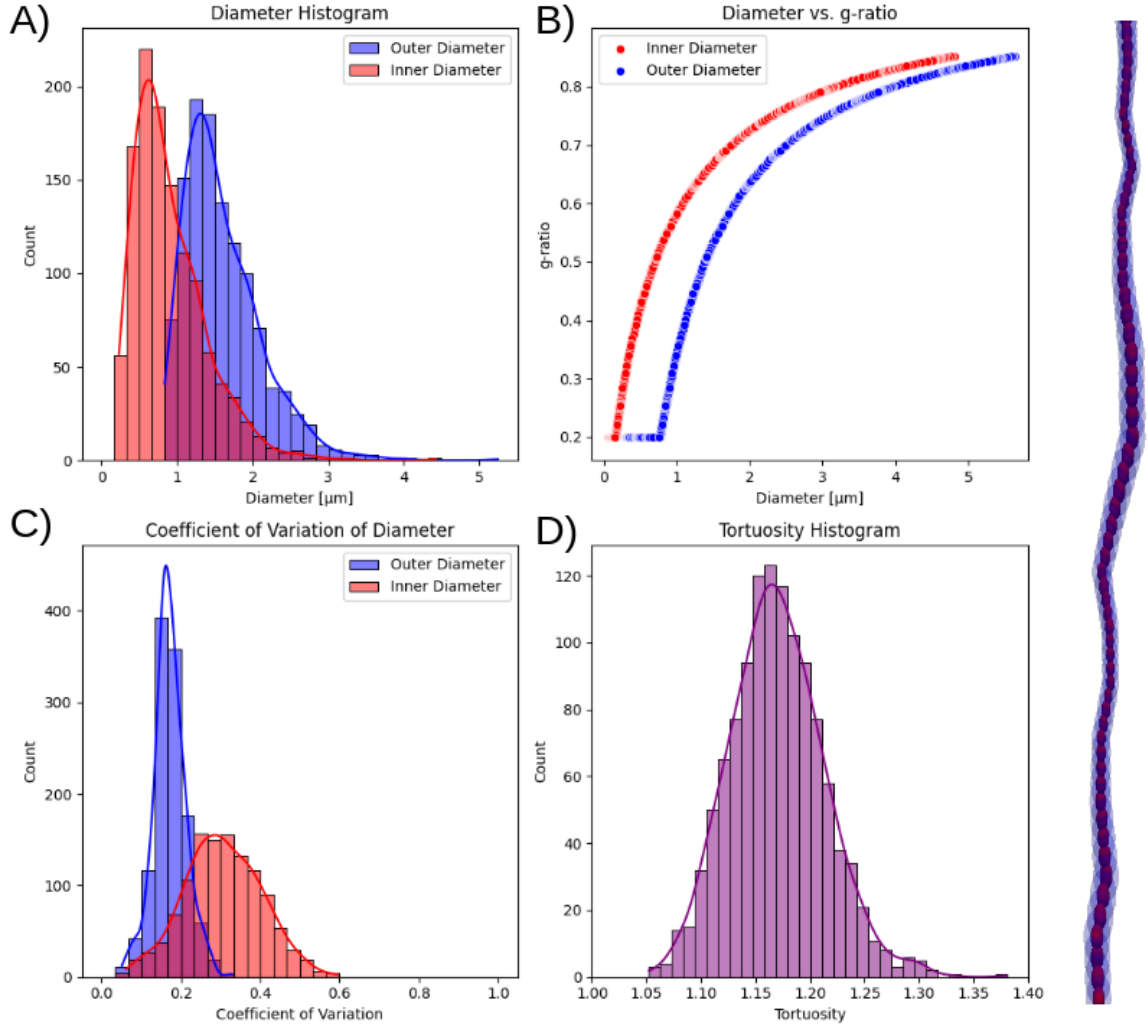


Figure 7: Several histograms showcasing the morphology of synthetic axons: A) Histogram of the inner (red) and outer (blue) diameters of approximately 1176 myelinated axons within a voxel of  $100 \mu m$  in length. The Gamma distribution parameters (for inner radius distribution) were  $\alpha = 4$  and  $\beta = 0.12$ . B) Histogram of the g-ratios of these axons with respect to the inner and outer diameter. This follows the log-linear relationship fit from [35]. C) Histogram of the Coefficients of Variation for inner and outer diameters, resulting from a beading amplitude of 0.3. D) Histogram of tortuosities, obtained using  $\epsilon = 0.4$ .

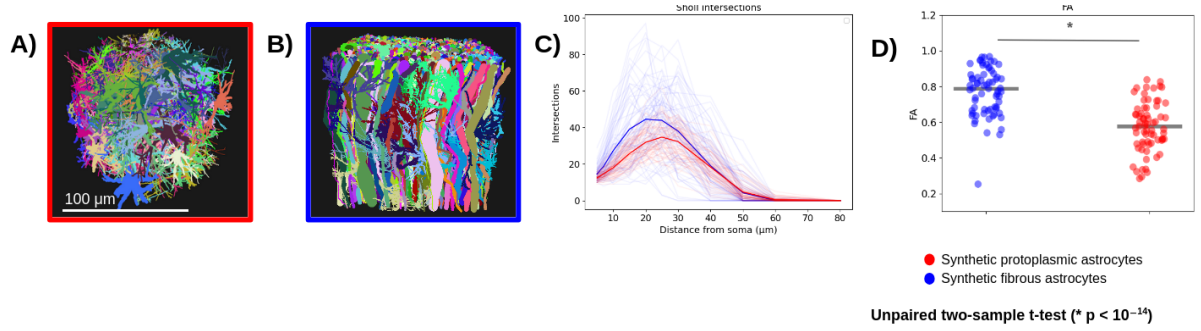


Figure 8: A) Illustration of synthetic astrocytes grown in empty space. The voxel is  $100 \mu\text{m}$  in length. B) Illustration of synthetic astrocytes grown in a voxel already filled with axons ( $f = 30\%$ ). C) Sholl analysis of the two kinds of astrocytes, in red for those shown in A and in blue for those in B. D) Differences in FA values between the two groups, with each scattered point referring to a specific astrocyte. Statistical significance was obtained using the unpaired two-sample t-test.

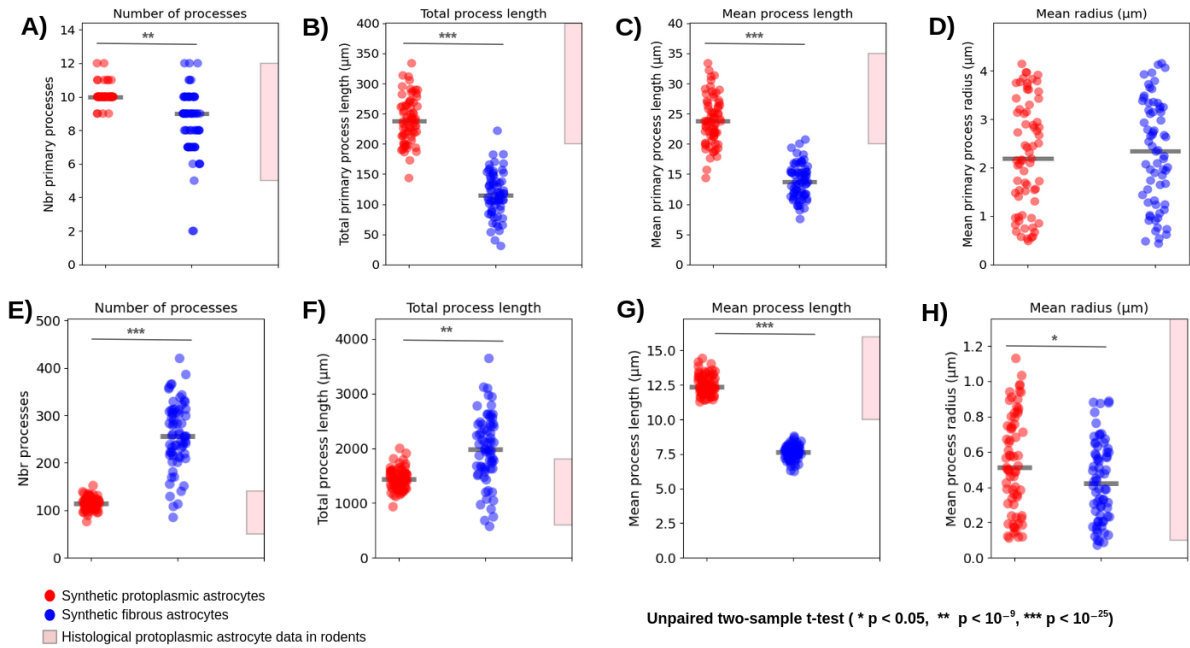


Figure 9: Number of processes, total process length, mean process length and mean process radius shown when taking into account A) only the primary processes, B) all processes. This is shown for synthetic protoplasmic and fibrous astrocytes. Each scattered point refers to an astrocyte. The pink boxes show the typical ranges for each morphological feature in histological findings for protoplasmic astrocytes in rodent GM [41, 42, 44, 46, 47, 55]. Statistical significance was obtained using the unpaired two-sample t-test.

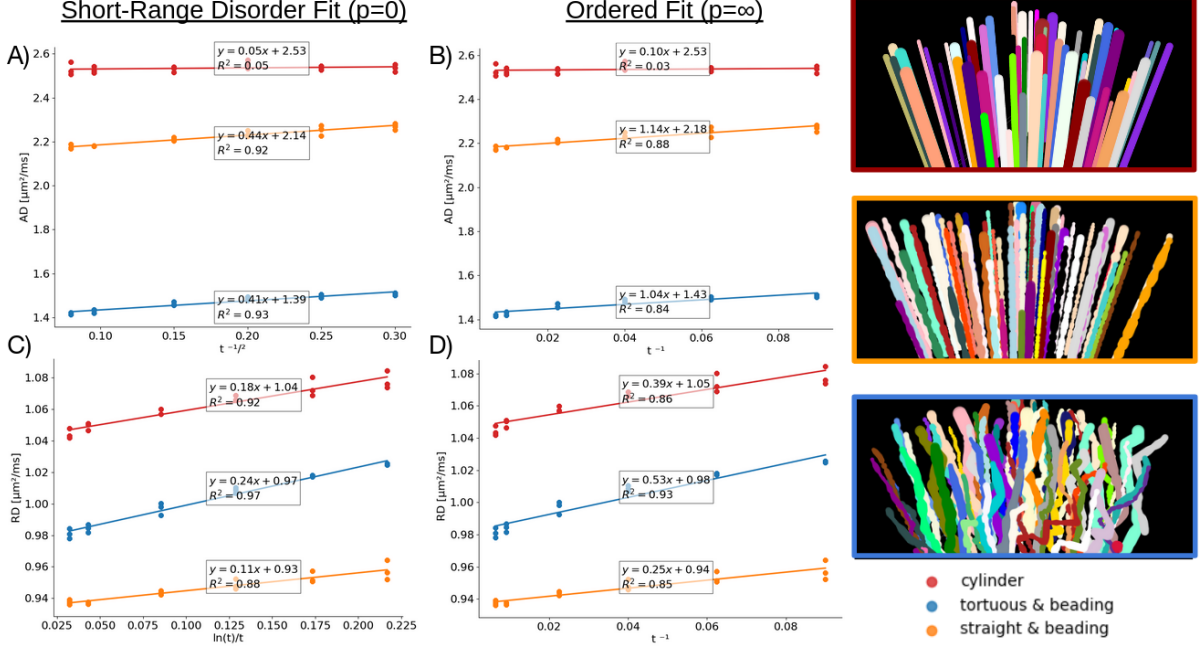


Figure 10: Time dependence for diffusion within the IAS along the axons (A and B) and within the EAS transverse to the axons (C and D). The time dependence is shown as a function of  $1/\sqrt{t}$  in A,  $\ln(t)/t$  in C and  $1/t$  in B and D. The results are shown for 3 different substrates: (i) with axons modelled as cylinders in red, (ii) with axons modelled as straight but with beading in orange and (iii) with axons modelled as tortuous and beading in blue. In the presence of short-range disorder, a better fit (given by  $R^2$ ) is found in A compared to B, and in C compared to D. Three repetitions were done per substrate. The solid lines represent linear fits to data.

Fitting the time dependence of the RD in EAS with  $\ln(t)/t$  consistently produced  $R^2$  values exceeding 0.88 across all substrates (Figure 10C), which were somewhat higher than the fits for  $1/t$  time dependence (Figure 10D). This indicates short-range disorder class characteristics for all substrates in the radial direction, including the substrate with cylinders. Among the substrates, the highest  $R^2$  value of 0.97 was observed for the tortuous beaded axons, indicating that this substrate shows the most prominent short-range disorder characteristic. The estimated  $D_\infty$  values for cylinders, straight axons, and tortuous axons were 1.04, 0.93, and  $0.97 \mu\text{m}^2/\text{ms}$ , respectively.

## 6 Discussion

We present CATERPillar, a new tool for generating numerical WM substrates that is inspired by the use of overlapping spheres (as in MEDUSA [26]) and of natural fibre growth (from CONFIG [27]). Unlike earlier tools focused mainly on axons, CATERPillar realistically models both axons and glial cells based on histological data. It supports the creation of beaded, tortuous, and dispersed axons (with or without myelin) and multiple glial populations, making it suitable for simulating physiologically relevant conditions and validating microstructural models. Voxel sizes ranged from 100–150  $\mu\text{m}$  in this study, though bigger voxels may be obtained at the cost of a longer run-time. The current packing limit is approximately 75%, and the maximum dispersion of fibres is  $c_2 = 0.75$  when only one fibre population is employed. Three crossing populations can achieve a  $c_2$  value for dispersion as low as 1/3.

### 6.1 Run-time

Creating numerical substrates of 150  $\mu\text{m}$  in length and with 50 % packing took approximately 2 hours. For substrates of 70% packing, this run-time reached 12 hours. A comparison with the CACTUS tool remained difficult, as the run-time published was for a 90% packing in very large voxels (500  $\mu\text{m}$ ). Compared to CONFIG and WMG, CATERPillar showed an improvement as they required approximately 6 hours and 31 hours, respectively, to generate a voxel of about 50  $\mu\text{m}$  in length. However, CATERPillar

remained much slower than the GPU-based MEDUSA, which could generate 100  $\mu\text{m}$  voxels in just 56 seconds. This is partly attributable to our use of a growth-based model instead of repulsion-based packing, which is more computationally efficient and allows for denser arrangements. Additionally, the use of CPU significantly slows substrate generation compared to GPU-based approaches. Future releases aim to improve runtime by implementing CATERPillar to be run on GPUs.

## 6.2 Realism of Synthetic Axons

To validate axonal realism, we compared diameter distribution, g-ratio scaling, beading and tortuosity with distribution obtained from histological data [35]. We found strong agreements that supports the biological accuracy of the generated axons. To obtain such accuracy, we used  $\epsilon = 0.4$  for tortuosity, beading amplitude = 0.3 and the Gamma distribution parameters (for inner radius distribution) were  $\alpha = 4$  and  $\beta = 0.12$ . If the user wishes to grow non-myelinated axons, we would advise a Gamma distribution with parameters  $\alpha = 4$  and  $\beta = 0.25$ , as this should be close to the corresponding outer radius distribution. Further adjustments can be made by the user if a specific WM region is targeted for example, and the axons are known to have different properties there. Beyond morphology, we evaluated the time dependence of diffusivity metrics in substrates with unmyelinated axons, a key feature of realistic and non-Gaussian diffusion. A decline in apparent diffusivity over time reflected structural complexity and restriction [56], which was especially expected in the most complex voxel containing beaded and tortuous axons. Since real WM tissue exhibits short-range disorder, theory predicts that AD in the IAS should scale with  $1/\sqrt{t}$  and RD in the EAS with  $\ln(t)/t$  [29, 57, 58]. Stronger dependence on  $1/t$  than these models would suggest periodicity within the structure, which is not realistic.

Our findings confirm that IAS diffusivity along the axons followed a  $1/\sqrt{t}$  scaling in both configurations where axons exhibit beading (whether tortuous or not), as opposed to cylindrical axons. We observed a similar restriction strength in both substrates that featured beaded axons (straight and tortuous), which suggests that beading is the primary contributor to the restriction strength as previously suggested [23]. Both  $D_\infty$  and the restriction strength obtained from our fit closely match values previously reported in segmented axons from EM and in vivo measurements in the brain’s corpus callosum [23]. For the EAS diffusivity perpendicular to the axons, the superior fit to  $\ln(t)/t$  as opposed to  $1/t$  across all substrates supported, once again, the presence of short-range disorder. Notably, this behaviour was observed even in substrate with cylindrical axons. In contrast to the IAS parallel diffusivity, such non-Gaussian diffusion within the EAS is expected due to the irregular spatial arrangement of axons, regardless of the complexity of the axonal morphology. The reduced  $D_\infty$  in the EAS observed in the substrate composed of straight beaded axons, relative to the tortuous and beaded configuration was surprising. This may be explained by the presence of small pockets or voids within the straight-beading architecture, which are features that may be less prevalent in the tortuous counterpart.

## 6.3 Realism of Synthetic Astrocytes

Astrocytic morphology was assessed in two conditions: substrates containing only astrocytes, and substrates where astrocytes grew in the presence of axons. We hypothesised that astrocytes grown in axon-rich environments will adopt a more elongated morphology compared to those generated in empty space, leading to higher FA values. This elongation would reflect the typical morphology of fibrous astrocytes found in real WM tissue, in contrast to the more isotropic shape of protoplasmic astrocytes commonly located in GM [45, 59]. Also, we expected morphological characteristics of our synthetic astrocytes to align with histological observations from rodent protoplasmic astrocytes, as their growth parameters were defined based on such data.

As anticipated, the FA of synthetic fibrous astrocytes was significantly higher than that of protoplasmic astrocytes. This confirms that, through simple rules conducting collision avoidance during process growth, we were successful in making our astrocytes adaptive to their morphology to their environment. Regarding the additional morphological characteristics, those obtained from synthetic protoplasmic astrocytes fell within the ranges reported in previous histological studies of rodent protoplasmic astrocytes, confirming that they exhibited realistic arbour complexity. This held for both user-defined features and those that were not explicitly controlled. The morphological characteristics of fibrous astrocytes differed from those of protoplasmic astrocytes, and were sometimes slightly outside of the expected range for histological protoplasmic astrocytes. These differences reflected the shorter and more processes of fibrous astrocytes, likely due to growth being constrained by frequent collisions with surrounding axons. In real tissue, fibrous astrocytes are known to have a larger size, less branched and straighter processes,

and fewer fine processes than protoplasmic astrocytes, which contrasts with our results. However, it is important to note that the same set of parameters were applied for both substrates to specifically demonstrate the impact of the environment on astrocyte morphology. Also, these parameters were set to specific values to match histological findings on protoplasmic astrocytes in rodents. To properly model fibrous astrocytes, the  $L_{process}$  would need to be much higher than the one used in this study. Keeping  $f_{soma}$  and  $f_{process}$  constant, the increase of  $L_{process}$  should reduce the need for branching, and thus decrease the number of processes.

## 7 Limitations

The main limitation of the CATERPillar tool lies in its achievable axonal volume fraction, which reaches approximately 75%, notably lower than the  $\sim 90\%$  reported with the CACTUS tool [28]. Unlike CACTUS, which uses an optimisation algorithm to resolve initial fibre overlaps, CATERPillar models axonal growth by extending fibres from an initial point toward an attractor, aiming to emulate natural development as previously done in CONFIG. While this biologically inspired approach may enhance realism, it also increases computational demands and inherently limits the maximum packing density. Nonetheless, a packing fraction of 75% is likely a realistic estimate, aligning with histological observations that typically report around 80% intracellular and 20% extracellular space [60]. It is worth noting, however, that chemical fixation procedures tend to shrink the extracellular space, potentially leading to an overestimation of intracellular volume fractions in histological data [61].

Additionally, in terms of axonal morphology, it has been noted that axonal cross-sections are not always perfectly round but can be elliptical in shape. A study using EM-based segmentation reported an average eccentricity of 0.72 for myelinated axons in the contralateral corpus callosum [24], indicating a notably elongated cross-sectional shape, as a circle would have an eccentricity of 0. While this elliptical cross-section was accounted for in the CACTUS [28], CONFIG [27] and WMG tools [5], it had not been explicitly incorporated into CATERPillar. Nonetheless, the cross-section of our substrates in the x-y plane seemed to naturally adopt elliptical shapes due to the tortuosity of the axons, but the resulting eccentricity was not assessed. The effective signature of elliptical cross sections on the dMRI signal remains to be clarified.

## 8 Conclusions and Future Works

The CATERPillar framework facilitates the generation of WM numerical substrates, incorporating axons, myelin, and glial cells. It enables efficient fibre growth toward predefined attractors, allowing multiple axons to develop simultaneously. Additionally, CATERPillar provides precise control over key morphological parameters, such as tortuosity and axonal beading. This tool can be used to (a) develop new acquisition schemes that sensitise the MRI signal to unique tissue microstructure features, (b) test the accuracy of a broad range of analytical models, and (c) build a set of substrates to train machine learning models on. Through this study, we have demonstrated that the generated substrates accurately replicate axonal and astrocytic morphological characteristics observed in histological studies, reinforcing the tool’s capability for realistic WM modelling. The resulting morphology yielded time-dependent diffusivities consistent with experimental observations of 1D and 2D structural disorder. Hence, we will provide an open-source software for creating realistic synthetic WM substrates, which will help advancing the capabilities of dMRI to quantify WM microstructure in vivo.

In future work, we aim to evaluate the performance of the Standard Model [10] on CATERPillar-generated substrates with progressively increasing structural complexity and biological realism. This will allow us to systematically investigate how factors such as beading and tortuosity, membrane permeability, orientation dispersion, fibre density, and astrocyte density influence the model’s predictive accuracy.

Furthermore, we plan to develop and train a supervised machine learning framework to infer microstructural parameters directly from diffusion signals. Because this approach does not rely on the conventional simplifying assumptions inherent to analytical models, it can learn complex relationships within the data and may ultimately yield more accurate and robust parameter estimates.

## 9 Declaration of Competing Interest

The authors declare no competing interest.



## **10 Code availability statement**

The GitHub code and GUI for CATERPillar will be made public upon publication of the manuscript.

## **11 Author Contribution**

Conceptualization: I.J, J.N.D; Methodology : I.J, J.N.D, M.C, M.B; Validation : J.N.D, M.B; Formal analysis: J.N.D; Writing-original draft: J.N.D, M.B; Writing - Review & Editing: I.J, J.N.D, M.B, J.P, I.d.R, J.R.P; Visualisation: J.N.D, M.B; Supervision: I.J.;

## **12 Acknowledgements**

This work was supported by the Swiss Secretariat for Research and Innovation (SERI) under an ERC Starting Grant award ‘FIREPATH’ MB22.00032.

## 13 Supplementary Material

### 13.1 Overlapping distance between unit spheres

To evaluate the effect of the spacing between overlapping spheres on diffusivity metrics, we compared synthetic DWI signals from straight axons modelled with varying sphere distances to those obtained using a perfect cylindrical representation. The objective was to identify the optimal  $F$  value that minimises discrepancies from the cylindrical model while balancing computational efficiency, as runtime increases with higher  $F$  values. To achieve this, we examined results for  $F = 1, 2, 4$ , and  $8$ .

Substrates with dimensions  $(150\ \mu\text{m})^3$  containing straight axons were generated, one for each specified value of  $F$ , plus an additional substrate where axons were represented as cylinders. The IAS volume fraction was consistently maintained at 50%. Simulations featured  $10^5$  random walker trajectories, with nine independent runs for each substrate type. A step size of  $0.1\ \mu\text{m}$  was employed, while IAS and EAS diffusivities were set at  $2\ \mu\text{m}^2/\text{ms}$ . The selected PGSE sequence used a wide pulse, defined by  $\Delta = 55.5\ \text{ms}$  and  $\delta = 16.5\ \text{ms}$ . DWIs were acquired along the x, y, and z axes, and apparent diffusion was calculated using the formula  $-\frac{\log(DWI_2/DWI_1)}{b_2 - b_1}$ , with  $b_1 = 0$  and  $b_2 = 1000\ \text{ms}/\mu\text{m}^2$ . RD was determined by measuring the apparent diffusion coefficient in the transverse direction to the axons (y-axis), while AD was calculated based on the apparent diffusion along the axonal orientation (z-axis). An unpaired two-sample t-test was applied to measure significant differences between the diffusivity metrics obtained with different values of  $F$ .

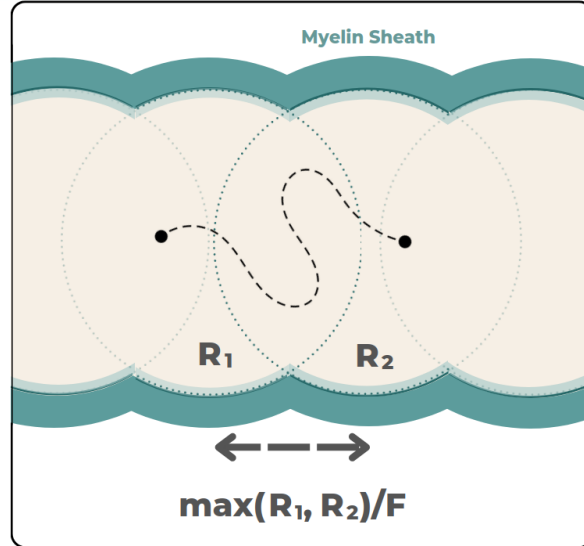


Figure S1: Illustration depicting how  $F$  is related to the distance between consecutive spheres and their inner radius.

The degree of overlap between the spheres, as expected, influenced both radial and axial diffusivity in the IAS and the EAS. As shown in Figure S2B, decreasing the distance between the spheres (or increasing the value of  $F$ ) caused the diffusivity values to more closely resemble those measured in a cylindrical substrate. In other words, higher  $F$  values reduced the deviation from the idealised diffusivity associated with cylinders. However, this improvement in accuracy came with a trade-off: as illustrated in Figure S2C, increasing  $F$  led to a steep rise in computational run-time, following an exponential trend.

The selected  $F$  value should therefore balance computational efficiency with reliable diffusivity results. Based on the data,  $F = 4$  emerged as a suitable compromise. With this choice, the computed diffusivity was sufficiently close to that of cylinders, while keeping the run-time manageable. Consequently, a value of  $F = 4$  was adopted for the analyses on the grown substrates.

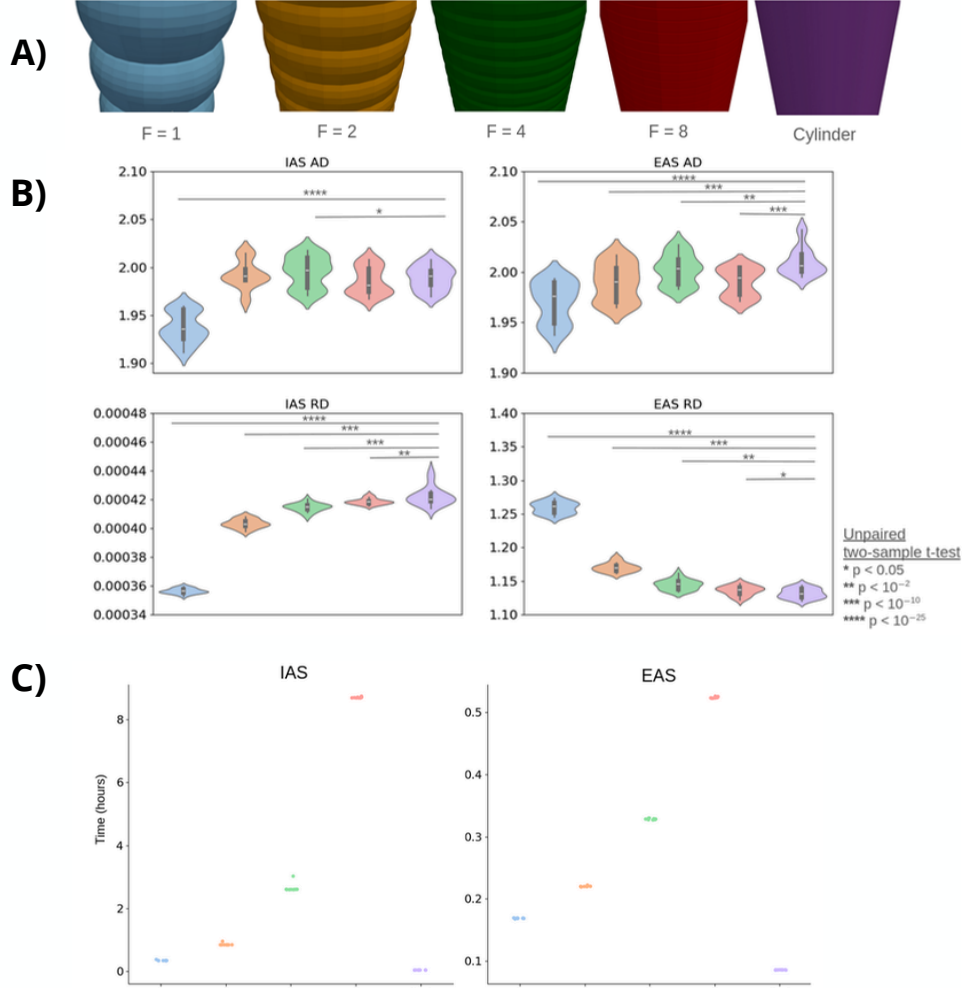


Figure S2: (A) Illustration of axonal segments demonstrating the influence of the factor  $F$  on morphology. The spacing between two spheres is determined by the maximum radius divided by  $F$ . (B) DKI-derived diffusivity estimates for the IAS and EAS across different  $F$  values. A trend of convergence is observed as  $F$  increases, with the highest  $F$  value yielding diffusivity estimates closest to those obtained from substrates where axons are modelled as cylinders. (C) Computational run-time for each type of substrate, showing an increasing trend with  $F$ , highlighting the trade-off between morphological accuracy and computational efficiency.

## 13.2 Effective medium theory

The dMRI signal represents the collective contribution of a vast number of diffusing spins. However, due to the macroscopic scale of imaging voxels, much of the fine-scale structural detail is averaged out and lost. Despite this, certain microstructural features persist in the dMRI signal and can be robustly quantified.

Effective medium theory [29] models tissue heterogeneity when structural variations occur at spatial scales comparable to the diffusion length of water that is much smaller than the imaging resolution. Instead of resolving each microstructural feature, it treats the complex tissue as a statistically homogeneous medium, where macroscopic observables are influenced by microscopic variability. This influence is captured through correlation functions that describe how structural fluctuations within the medium affect the overall diffusion signal measured at the voxel level.

Different media can be categorised into structural universality classes, each characterised by a distinct structural exponent  $p$  [29]. One of these classes is called “Order” and corresponds to periodic arrangements with  $p = \infty$ . Another class, which is the most commonly observed, is called “Short-Range Disorder” and is associated with  $p = 0$ . The “Hyperuniform Disorder” class represents systems with  $p > 0$ , while the “Strong Disorder” class corresponds to  $p < 0$ . The exponent  $p$  provides a quantitative measure of the spatial correlations in the underlying microstructural heterogeneity for each class. It can be derived from the two-point correlator function  $\Gamma(\mathbf{q})$ , which is the Fourier transform of the binary mask representing the medium. When the wave vector  $\mathbf{q}$  is sufficiently low ( $q \rightarrow 0$ ), then  $\Gamma(\mathbf{q})$  behaves like  $q^p$  [57]. The exponent  $p$  can thus be determined by how  $\Gamma(\mathbf{q})$  behaves in the low- $q$  regime, and enables the grouping of differently looking media into distinct structural universality classes. Histological data analysis suggests that the most appropriate universality class for describing tissue structure is the short-range disorder [16, 23, 57, 58]. Since this class is associated with  $p = 0$ , we expect  $\Gamma(\mathbf{q})$  to have a finite plateau in the low- $q$  regime. This indicates the absence of long-range correlations beyond the disorder correlation length  $l_c \sim \frac{1}{q_c}$ . The parameter  $q_c$  defines the extent of this low- $q$  plateau [57]. Physically,  $l_c$  represents the characteristic distance from a given restriction beyond which positional information about other restrictions is lost, effectively erasing any “memory” of their locations [58].

Diffusion in media belonging to different structural universality classes exhibits distinct time-dependent behaviour. In particular, for short-range disordered media, the instantaneous diffusivity  $D(t)$  along the principal axis of axons in the intra-axonal space (IAS), under one-dimensional diffusion ( $d = 1$ ), is expected to follow the relationship:

$$D(t) = \begin{cases} D_\infty + c_1 t^{-\theta}, & \theta < 1 \\ D_\infty + c_2 \frac{\ln(t/t_c)}{t}, & \theta = 1 \\ D_\infty + \frac{c_3}{t}, & \theta > 1 \end{cases} \quad \text{with} \quad \theta = \frac{p+d}{2}, t \gg t_c = \frac{l_c^2}{2D_\infty} \quad (2)$$

where  $D_\infty$  is the apparent diffusion when  $t \rightarrow \infty$ ,  $c_1$ ,  $c_2$  and  $c_3$  are constant values,  $t$  is the diffusion time,  $p$  is the structural exponent,  $t_c$  represents the time to diffuse across the correlation length  $l_c$  and  $d$  is the number of dimensions. This applies to a PGSE acquisition in the narrow pulse limit, where the gradient duration  $\delta$  is very short. However, additional corrections are necessary to account for the wide pulse limit, in which  $\delta$  is of the order of gradient separation  $\Delta$  [16]. In short-range disorder ( $p = 0$ ), the value of  $\theta$  is 1/2 for 1D diffusion along the primary orientation of the axons in the IAS. The equation then becomes :

$$D(t) = D_\infty + c_1 t^{-1/2} \quad (3)$$

When studying the two-dimensional EAS diffusivity transverse to the main orientation of the axons, the equation becomes :

$$D(t) = D_\infty + c_2 \frac{\ln(t/t_c)}{t} \quad (4)$$

When the structural constraints exhibit periodicity ( $p = \infty$ ),  $D(t)$  follows a characteristic scaling behaviour, regardless of the number of spatial dimensions  $d$ :

$$D(t) = D_\infty + c_3 \frac{1}{t} \quad (5)$$

If no heterogeneity is present on the scale of diffusion length, diffusion in each compartment remains Gaussian and diffusivity does not depend on diffusion time. This is typically observed along the length

of a perfect cylinder. Previous research have demonstrated the presence of a time-dependent behaviour in WM tissue across various settings. Monte Carlo simulations conducted on axons segmented from EM data have confirmed this effect [16, 23]. Additionally, similar findings have been reported in experimental phantoms composed of parallel fibres designed to mimic axons [57], as well as in *in vivo* studies [58, 62]. These findings align with the effective medium theory and prove that the diffusion in the IAS and EAS is indeed non-Gaussian, as opposed to assumptions made by the SM. This behaviour is thought to be due to restrictions experienced by water molecules within the biological tissue, which are induced by axonal beading and undulations in the longitudinal direction, and the random packing geometry of fibres within a bundle in the transverse direction [58].

### 13.3 Structural correlations within numerical substrates

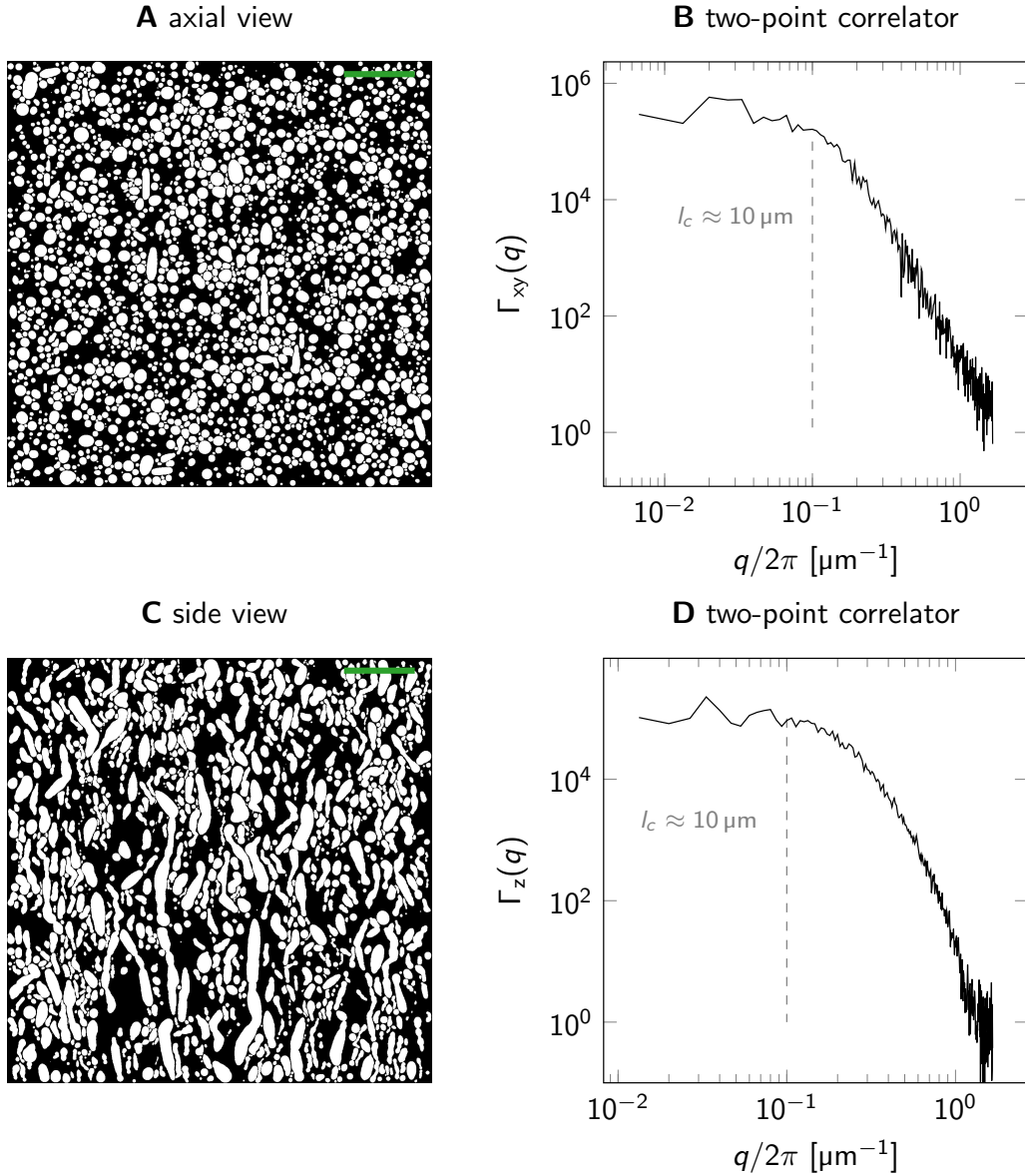


Figure S3: A) Axial view illustrating synthetic axons within a voxel with a volume fraction  $f = 0.45$ . B) The two-point correlation function  $\Gamma_{xy}$  in the transverse plane of the generated voxel. Below a spatial frequency of  $q = 2\pi/10 \mu\text{m}$ , a plateau is observed, which corresponds to short-range disorder for lengths longer than the correlation length  $l_c \approx 10 \mu\text{m}$ . C) Lateral (side) view of the same voxel. D) The two-point correlation function  $\Gamma_z$  computed along the z-axis, i.e., the primary direction of the axons. As in B, short-range disorder is observed with a similar correlation length of  $l_c \approx 10 \mu\text{m}$ . The scale bar in A and C corresponds to  $25 \mu\text{m}$ .

Figure 10 analyses the time-dependence of dMRI signals in our numerical substrates, confirming their classification as short-range disordered media. To further substantiate this, we also compute the two-point correlation function for the beaded and tortuous substrate used in that analysis. Prior studies have shown that this approach can reveal short-range disorder in axonal organisation without relying on MC simulations [23, 29]. As such, it provides complementary evidence that our substrates exhibit the expected disordered structure. In this analysis, it is indicated by a plateau in the two-point correlation function at low spatial frequencies, reflecting the absence of spatial correlations beyond a finite correlation length  $l_c$ .

We generated a 3D binary mask  $\varrho$  that samples the volume of the numerical substrate of interest at an isotropic resolution of  $0.3\mu\text{m}$ . The two-point correlator was calculated from the Fourier transform of the mask using the Wiener-Khinchin theorem  $\Gamma(\mathbf{q}) = |\varrho(\mathbf{q})|^2/V$ , where  $V$  is the substrate volume. We estimated the two-point correlation function  $\Gamma$  in the  $x$ - $y$  plane by averaging  $\Gamma$  along both axes to obtain  $\Gamma_{xy}$  (Figure S3A and B), which reflects the structural organisation perpendicular to the axons and along the  $z$ -axis (Figure S3C and D), representing the primary axonal orientation. Both the two-point correlation function in the plane perpendicular to the axon and along the axons indeed showed a plateau for frequencies below  $q_c \approx 1/10\mu\text{m}$ . This complementary analysis therefore confirms the expected short-range disorder, with correlation lengths exceeding  $l_c \approx 10\mu\text{m}$ .

## References

- [1] Margaret Flanagan, Joshua A. Sonnen, Christopher Dirk Keene, Robert F. Hevner, and Thomas J. Montine. Chapter 29 - molecular basis of diseases of the nervous system. In William B. Coleman and Gregory J. Tsongalis, editors, *Molecular Pathology (Second Edition)*, pages 651–690. Academic Press, second edition edition, 2018.
- [2] Jiaojiao Zhang, Rosario Cortese, Nicola De Stefano, and Annamaria Giorgio. Structural and functional connectivity substrates of cognitive impairment in multiple sclerosis. *Frontiers in Neurology*, 12:671894, July 8 2021.
- [3] David Attwell, Alastair M. Buchan, Serge Charpak, Martin Lauritzen, Brian A. Macvicar, and Eric A. Newman. Glial and neuronal control of brain blood flow. *Nature*, 468(7321):232–243, 2010.
- [4] Dale Purves, George J Augustine, David Fitzpatrick, Lawrence C Katz, Anthony-Samuel LaMantia, James O McNamara, and S Mark Williams. Neuroglial cells. *Neuroscience*, 2001.
- [5] Sidsel Winther, Oscar Peulicke, Mariam Andersson, Hans M. Kjer, Jakob A. Bærentzen, and Tim B. Dyrby. Exploring white matter dynamics and morphology through interactive numerical phantoms: the white matter generator. *Frontiers in Neuroinformatics*, 18, 2024.
- [6] Mariam Andersson, Hans Martin Kjer, Jonathan Rafael-Patino, Alexandra Pacureanu, Bente Pakkenberg, Jean-Philippe Thiran, Maurice Ptito, Martin Bech, Anders Bjorholm Dahl, Vedrana Andersen Dahl, and Tim B. Dyrby. Axon morphology is modulated by the local environment and impacts the noninvasive investigation of its structure–function relationship. *Proceedings of the National Academy of Sciences*, 117(52):33649–33659, 2020.
- [7] Daniel C. Alexander, Tim B. Dyrby, Markus Nilsson, and Hui Zhang. Imaging brain microstructure with diffusion mri: practicality and applications. *NMR in Biomedicine*, 32(4):e3841, 2019. e3841 NBM-16-0261.R1.
- [8] Valerij G. Kiselev. Fundamentals of diffusion mri physics. *NMR in Biomedicine*, 30(3):e3602, 2017. e3602 nbm.3602.
- [9] Ileana O. Jelescu, Marco Palombo, Francesca Bagnato, and Kurt G. Schilling. Challenges for biophysical modeling of microstructure. *Journal of Neuroscience Methods*, 344:108861, 2020.
- [10] Dmitry S Novikov, Els Fieremans, Sune N Jespersen, and Valerij G Kiselev. Quantifying brain microstructure with diffusion mri: Theory and parameter estimation. *NMR in Biomedicine*, 32(4):e3998, 2019.
- [11] Greg J Stanisz, Graham A Wright, R Mark Henkelman, and Aaron Szafer. An analytical model of restricted diffusion in bovine optic nerve. *Magnetic Resonance in Medicine*, 37(1):103–111, 1997.
- [12] Els Fieremans and Hong-Hsi Lee. Physical and numerical phantoms for the validation of brain microstructural mri: A cookbook. *NeuroImage*, 182:39–61, 2018. Microstructural Imaging.
- [13] Chun-Hung Yeh, Benoît Schmitt, Denis Le Bihan, Jing-Rebecca Li-Schlittgen, Ching-Po Lin, and Cyril Poupon. Diffusion microscopist simulator: a general monte carlo simulation system for diffusion magnetic resonance imaging. *PloS one*, 8(10):e76626, 2013.
- [14] Matt G Hall and Daniel C Alexander. Convergence and parameter choice for monte-carlo simulations of diffusion mri. *IEEE transactions on medical imaging*, 28(9):1354–1364, 2009.
- [15] Jonathan Rafael-Patino, David Romascano, Alonso Ramirez-Manzanares, Erick Jorge Canales-Rodríguez, Gabriel Girard, and Jean-Philippe Thiran. Robust monte-carlo simulations in diffusion-mri: Effect of the substrate complexity and parameter choice on the reproducibility of results. *Frontiers in Neuroinformatics*, 14:8, 2020.
- [16] H. H. Lee, E. Fieremans, and D. S. Novikov. Realistic microstructure simulator (rms): Monte carlo simulations of diffusion in three-dimensional cell segmentations of microscopy images. *Journal of Neuroscience Methods*, 350:109018, Feb 2021. Epub 2020 Dec 3.
- [17] Michiel Cottaar. Mcmrsimulator.jl, November 2022.



- [18] Matthew D. Budde and Joseph A. Frank. Neurite beading is sufficient to decrease the apparent diffusion coefficient after ischemic stroke. *Proceedings of the National Academy of Sciences*, 107(32):14472–14477, 2010.
- [19] Arthur P C Spencer, Jasmine Nguyen-Duc, Inès de Riedmatten, Filip Szczepankiewicz, and Ileana O Jelescu. Mapping grey and white matter activity in the human brain with isotropic adc-fmri. *bioRxiv*, 2024.
- [20] Chih-Liang Chin, Felix W Wehrli, Scott N Hwang, Masaya Takahashi, and David B Hackney. Bi-exponential diffusion attenuation in the rat spinal cord: computer simulations based on anatomic images of axonal architecture. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 47(3):455–460, 2002.
- [21] Khieu-Van Nguyen, Edwin Hernández-Garzón, and Julien Valette. Efficient gpu-based monte-carlo simulation of diffusion in real astrocytes reconstructed from confocal microscopy. *Journal of Magnetic Resonance*, 296:188–199, 2018.
- [22] Tianyou Xu, Sean Foxley, Michiel Kleinnijenhuis, Way Cherng Chen, and Karla L. Miller. The effect of realistic geometries on the susceptibility-weighted mr signal in white matter. *Magnetic Resonance in Medicine*, 79(1):489–500, 2018.
- [23] Heesoo H. Lee, Angelos Papaioannou, Sungheon L. Kim, Dmitry S. Novikov, and Els Fieremans. A time-dependent diffusion mri signature of axon caliber variations and beading. *Communications Biology*, 3(1):354, Jul 2020.
- [24] Ali Abdollahzadeh, Ilya Belevich, Eija Jokitalo, Jussi Tohka, and Alejandra Sierra. Deepacson automated segmentation of white matter in 3d electron microscopy. *Communications Biology*, 4:179, 2021.
- [25] Gregory T. Balls and Lawrence R. Frank. A simulation environment for diffusion weighted mr experiments in complex media. *Magnetic Resonance in Medicine*, 62(3):771–778, 2009.
- [26] Kévin Ginsburger, Felix Matuschke, Fabrice Poupon, Jean-François Mangin, Markus Axer, and Cyril Poupon. Medusa: A gpu-based tool to create realistic phantoms of the brain microstructure using tiny spheres. *NeuroImage*, 193:10–24, 2019.
- [27] Ross Callaghan, Daniel C. Alexander, Marco Palombo, and Hui Zhang. Config: Contextual fibre growth to generate realistic axonal packing for diffusion mri simulation. *NeuroImage*, 220:117107, 2020.
- [28] J. L. Villarreal-Haro, R. Gardier, E. J. Canales-Rodríguez, E. Fisch-Gomez, G. Girard, J. P. Thiran, and J. Rafael-Patiño. Cactus: a computational framework for generating realistic white matter microstructure substrates. *Frontiers in Neuroinformatics*, 17:1208073, 2023.
- [29] Dmitry S. Novikov, Jens H. Jensen, Joseph A. Helpert, and Els Fieremans. Revealing mesoscopic structural universality with diffusion. *Proceedings of the National Academy of Sciences*, 111(14):5088–5093, 2014.
- [30] Y. Assaf, T. Blumenfeld-Katzir, Y. Yovel, and P. J. Basser. Axciliber: a method for measuring axon diameter distribution from diffusion mri. *Magnetic Resonance in Medicine*, 59:1347–1354, 2008.
- [31] R. Shekhar, E. Fayyad, R. Yagel, and J. F. Cornhill. Octree-based decimation of marching cubes surfaces. In *Proceedings of Seventh Annual IEEE Visualization’96*, pages 335–342, San Francisco, CA, 1996. IEEE.
- [32] Quentin Avril, Valérie Gouranton, and Bruno Arnaldi. Dynamic adaptation of broad phase collision detection algorithms. In *ISVRI 2011 - IEEE International Symposium on Virtual Reality Innovations 2011, Proceedings*, pages 41–47, 2011.
- [33] Armen Stepanyants, Gábor Tamás, and Dmitri B Chklovskii. Class-specific features of neuronal wiring. *Neuron*, 43(2):251–259, 2004.
- [34] Claes-Henric Berthold, I Nilsson, and M Rydmark. Axon diameter and myelin sheath thickness in nerve fibres of the ventral spinal root of the seventh lumbar nerve of the adult and developing cat. *Journal of anatomy*, 136(Pt 3):483, 1983.

- [35] H. H. Lee, K. Yaros, J. Veraart, and et al. Along-axon diameter variation and axonal orientation dispersion revealed with 3d electron microscopy: implications for quantifying brain white matter microstructure with histology and diffusion mri. *Brain Struct Funct*, 224:1469–1488, 2019.
- [36] Nancy Ann Oberheim, Takahiro Takano, Xiaoning Han, Wei He, Jane H. C. Lin, Fushun Wang, Qiwu Xu, Jeffrey D. Wyatt, Webster Pilcher, Jeffrey G. Ojemann, Bruce R. Ransom, Steven A. Goldman, and Maiken Nedergaard. Uniquely hominid features of adult human astrocytes. *Journal of Neuroscience*, 29(10):3276–3287, 2009.
- [37] Leonid P. Savtchenko, Lise Bard, Thomas P. Jensen, James P. Reynolds, Igor Kraev, Nikolay Medvedev, Dagmar Minge, Denis E. Korzhhevskii, Valeriy A. Makarov, Alexei Verkhratsky, and Dmitri A. Rusakov. Disentangling astroglial physiology with a realistic cell model in silico. *Nature Communications*, 9(1):3554, 2018.
- [38] Alexander M. Reeves, Eiji Shigetomi, and Baljit S. Khakh. Bulk loading of calcium indicator dyes to study astrocyte physiology: key limitations and improvements using morphological maps. *Journal of Neuroscience*, 31(25):9353–9358, June 22 2011.
- [39] Wenyuan Xin, Kathryn E. Schuebel, Kwok Jair, and et al. Ventral midbrain astrocytes display unique physiological features and sensitivity to dopamine d2 receptor signaling. *Neuropsychopharmacology*, 44:344–355, 2019.
- [40] Alexei Verkhratsky, Nancy Bush, Maiken Nedergaard, and Arthur Butt. The special case of human astrocytes. *Neuroglia*, 1:4, 03 2018.
- [41] João Filipe Viana, João Luís Machado, Daniela Sofia Abreu, Alexandra Veiga, Sara Barsanti, Gabriela Tavares, Manuella Martins, Vanessa Morais Sardinha, Sónia Guerra-Gomes, Cátia Domingos, Alberto Pauletti, Jérôme Wahis, Chen Liu, Corrado Calì, Christian Henneberger, Matthew G. Holt, and João Filipe Oliveira. Astrocyte structural heterogeneity in the mouse hippocampus. *Glia*, 71(7):1667–1682, 2023.
- [42] Lin He, Shaojun Ma, Zhen Ding, Zihan Huang, Yifan Zhang, Chen Xi, Kun Zou, Qing Deng, Wenjian M. Huang, Qiang Guo, and Chun Huang. Inhibition of nfat5-dependent astrocyte swelling alleviates neuropathic pain. *Advanced Science*, 11(11):e2302916, March 2024. Epub 2024 Jan 9.
- [43] F. Althammer, H. C. Ferreira-Neto, M. Rubaharan, et al. Three-dimensional morphometric analysis reveals time-dependent structural changes in microglia and astrocytes in the central amygdala and hypothalamic paraventricular nucleus of heart failure rats. *Journal of Neuroinflammation*, 17(221), 2020.
- [44] Marcel Klein, Christian Lohr, and Damian Droste. Age-dependent heterogeneity of murine olfactory bulb astrocytes. *Frontiers in Aging Neuroscience*, 12, 06 2020.
- [45] Susanne Köhler, Ulrike Winkler, Tabea Junge, Kristina Lippmann, Jens Eilers, and Johannes Hirrlinger. Gray and white matter astrocytes differ in basal metabolism but respond similarly to neuronal activity. *Glia*, 71(2):229–244, 2023.
- [46] Sara Barsanti, João Filipe Viana, Alexandra Veiga, João Luís Machado, Daniela Sofia Abreu, José Duarte Dias, Susana Monteiro, Nuno A. Silva, Luísa Pinto, and João Filipe Oliveira. Assessing different histological preparations for reconstruction of astrocyte tridimensional structure. *Cells*, 13(11), 2024.
- [47] Charlie Aird-Rossiter, Hui Zhang, Daniel C. Alexander, Derek K. Jones, and Marco Palombo. Decoding gray matter: large-scale analysis of brain cell morphometry to inform microstructural modeling of diffusion mr signals, 2025.
- [48] Kelsey T. Baldwin, Keith K. Murai, and Baljit S. Khakh. Astrocyte morphology. *Trends in Cell Biology*, 34(7):547–565, July 2024.
- [49] Rémy Gardier, Juan Luis Villarreal Haro, Erick J. Canales-Rodríguez, Ileana O. Jelescu, Gabriel Girard, Jonathan Rafael-Patiño, and Jean-Philippe Thiran. Cellular exchange imaging (cexi): Evaluation of a diffusion model including water exchange in cells using numerical phantoms of permeable spheres. *Magnetic Resonance in Medicine*, 90(4):1625–1640, 2023.

- [50] Dmitry S. Novikov, Jelle Veraart, Ileana O. Jelescu, and Els Fieremans. Rotationally-invariant mapping of scalar and orientational metrics of neuronal microstructure with diffusion mri. *NeuroImage*, 174:518–538, 2018.
- [51] Santiago Coelho, Steven H. Baete, Gregory Lemberskiy, Benjamin Ades-Aron, Genevieve Barrol, Jelle Veraart, Dmitry S. Novikov, and Els Fieremans. Reproducibility of the standard model of diffusion in white matter on clinical mri systems. *NeuroImage*, 257:119290, 2022.
- [52] Benjamin Ades-Aron. Evaluation of the accuracy and precision of the diffusion parameter estimation with gibbs and noise removal pipeline. *NeuroImage*, 183:532–543, 2018.
- [53] David Romascano, Juan Rafael-Patino, Ileana Jelescu, Mirsad Barakovic, Ben Tim, Doyon Jean-Philippe, et al. Voxel size matters: big voxels are required to generate realistic extra-axonal DMRI signals from Monte Carlo simulations. In *Proceedings of the International Society for Magnetic Resonance in Medicine (ISMRM)*, pages 1–2, 2018.
- [54] Emmanuel Caruyer, Christophe Lenglet, Guillermo Sapiro, and Rachid Deriche. Design of multi-shell sampling schemes with uniform coverage in diffusion MRI. *Magnetic Resonance in Medicine*, 69(6):1534–1540, Jun 2013. Epub 2013 Apr 26.
- [55] G. Tavares, M. Martins, J.S. Correia, V.M. Sardinha, S. Guerra-Gomes, S.P. das Neves, F. Marques, N. Sousa, and J.F. Oliveira. Employing an open-source tool to assess astrocyte tridimensional structure. *Brain Struct Funct*, 222(4):1989–1999, May 2017. Epub 2016 Sep 30.
- [56] Ileana O. Jelescu, Alexandre de Skowronski, Françoise Geffroy, Marco Palombo, and Dmitry S. Novikov. Neurite exchange imaging (nexi): A minimal model of diffusion in gray matter with inter-compartment water exchange. *NeuroImage*, 256:119277, 2022.
- [57] Lauren M. Burcaw, Els Fieremans, and Dmitry S. Novikov. Mesoscopic structure of neuronal tracts from time-dependent diffusion. *NeuroImage*, 114:18–37, 2015.
- [58] Els Fieremans, Lauren M. Burcaw, Hong-Hsi Lee, Gregory Lemberskiy, Jelle Veraart, and Dmitry S. Novikov. In vivo observation and biophysical interpretation of time-dependent diffusion in human white matter. *NeuroImage*, 129:414–427, 2016.
- [59] Marco Palombo, Noam Shemesh, Itamar Ronen, and Julien Valette. Insights into brain microstructure from in vivo dw-mrs. *NeuroImage*, 182, 11 2017.
- [60] Eva Syková and Charles Nicholson. Diffusion in brain extracellular space. *Physiological Reviews*, 88(4):1277–1340, 2008. PMID: 18923183.
- [61] Nina Korogod, Carl C. H. Petersen, and Graham W. Knott. Ultrastructural analysis of adult mouse neocortex comparing aldehyde perfusion with cryo fixation. *eLife*, 4:e05793, August 11 2015.
- [62] Sune Nørhøj Jespersen, Jonas Lynge Olesen, Brian Hansen, and Noam Shemesh. Diffusion time dependence of microstructural parameters in fixed spinal cord. *NeuroImage*, 182:329–342, 2018. Microstructural Imaging.