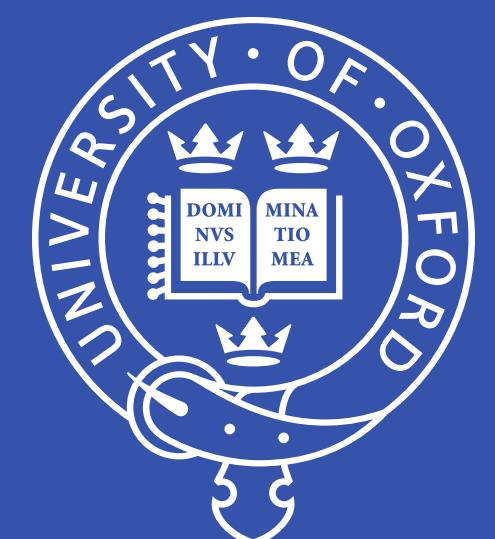


# White matter tractography and Human Brain Connections using GPUs

Moisés Hernández-Fernandez<sup>1</sup>, Istvan Reguly<sup>2,3</sup>, Mike Giles<sup>2</sup>, Stephen Smith<sup>1</sup> and Stamatis N. Sotiropoulos<sup>1</sup>

<sup>1</sup>Oxford Centre for Functional MRI of the Brain (FMRIB), University of Oxford, UK    <sup>2</sup>Oxford e-Research Centre, University of Oxford, UK

<sup>3</sup>Faculty of Information Technology and Bionics, Pazmany Peter Catholic University, Hungary

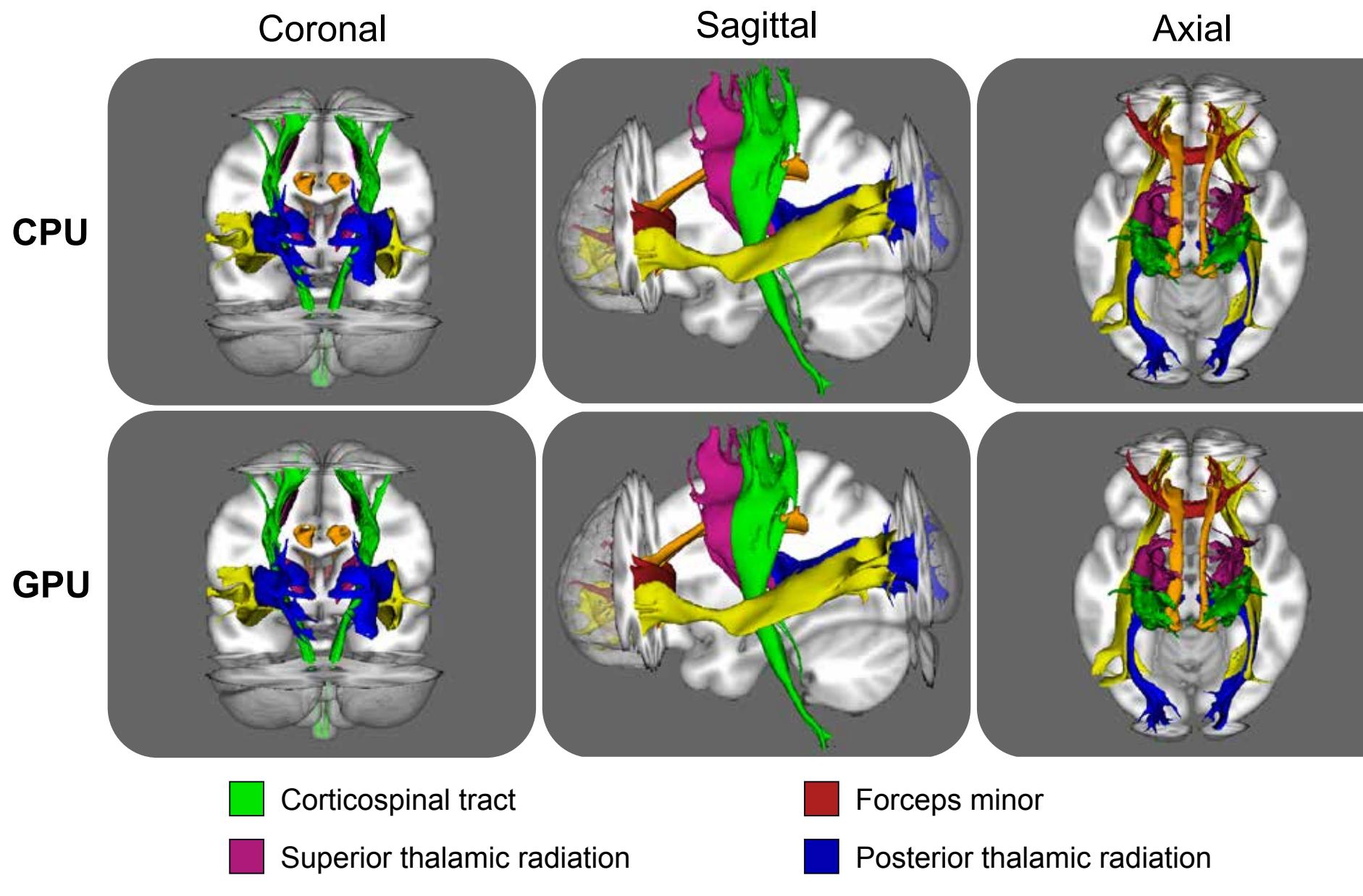


## Abstract

Diffusion MRI (dMRI) allows the study of brain microstructure and permits the estimation of long-range brain connections, non-invasively and in-vivo (tractography). Despite its great potential, the analysis includes the very time-consuming statistical modelling of four-dimensional datasets that restrict the potential and the clinical applicability of the technology. We present a novel parallel framework for analysis of dMRI data that allows accelerations of up to two orders of magnitude when comparing GPU performance with a single-threaded CPU implementation.

## 1. Introduction: Diffusion MRI and tractography

Diffusion MRI (dMRI) and tractography methods allow uniquely the in-vivo study of anatomical connections in the human brain. These connections allow information flow between remote areas and form the backbone of brain function.



Their study is crucial in gaining insight into the complex functional organisation of the brain that leads to various aspects of human behaviour.

The connections may also be affected by pathology and their study can shed light to disease aetiology. For instance, anatomical brain connectivity has been shown to be different in groups with Alzheimer's disease and multiple sclerosis. The representation of white matter pathways of a patient can also guide neurosurgeons in presurgical planning and intraoperative neuronavigation.

Understanding the human brain is one of the key scientific challenges of the 21<sup>st</sup> century. The potential of dMRI cannot be underestimated given its application in very large recent studies such as the Human Connectome Project (HCP) or the UK Biobank project.

## 4. Conclusions

- We have designed and implemented parallel algorithms on Nvidia GPUs to estimate the distribution of fibre orientations in the white matter from diffusion MR images and to perform probabilistic tractography. We have validated our GPU parallel algorithms comparing them with the CPU algorithms and we did not find significant differences between the results on the GPU and the CPU.
- The fibre orientation estimation algorithm has been accelerated by at least two orders of magnitude (Results published in Hernández, Moisés, et al. "Accelerating fibre orientation estimation from diffusion weighted magnetic resonance imaging using GPUs." (2013): e61892.), when comparing a single GPU with the respective sequential single-core CPU version, as well as when comparing multi-GPU with multi-CPU implementations. For the probabilistic tractography algorithm we show speed-up factors of up to 180x in a single GPU compared to its sequential single-core CPU counterpart version.
- Achieving such high accelerations is tremendously beneficial, especially in very large recent studies such as the Human Connectome Project, where comprehensive maps of brain anatomical connectivity of unprecedented quality are being generated. These accelerations can change the perspective of what is feasible and allow us to extract the most out of the latest state-of-the-art datasets.

## 2. Methods

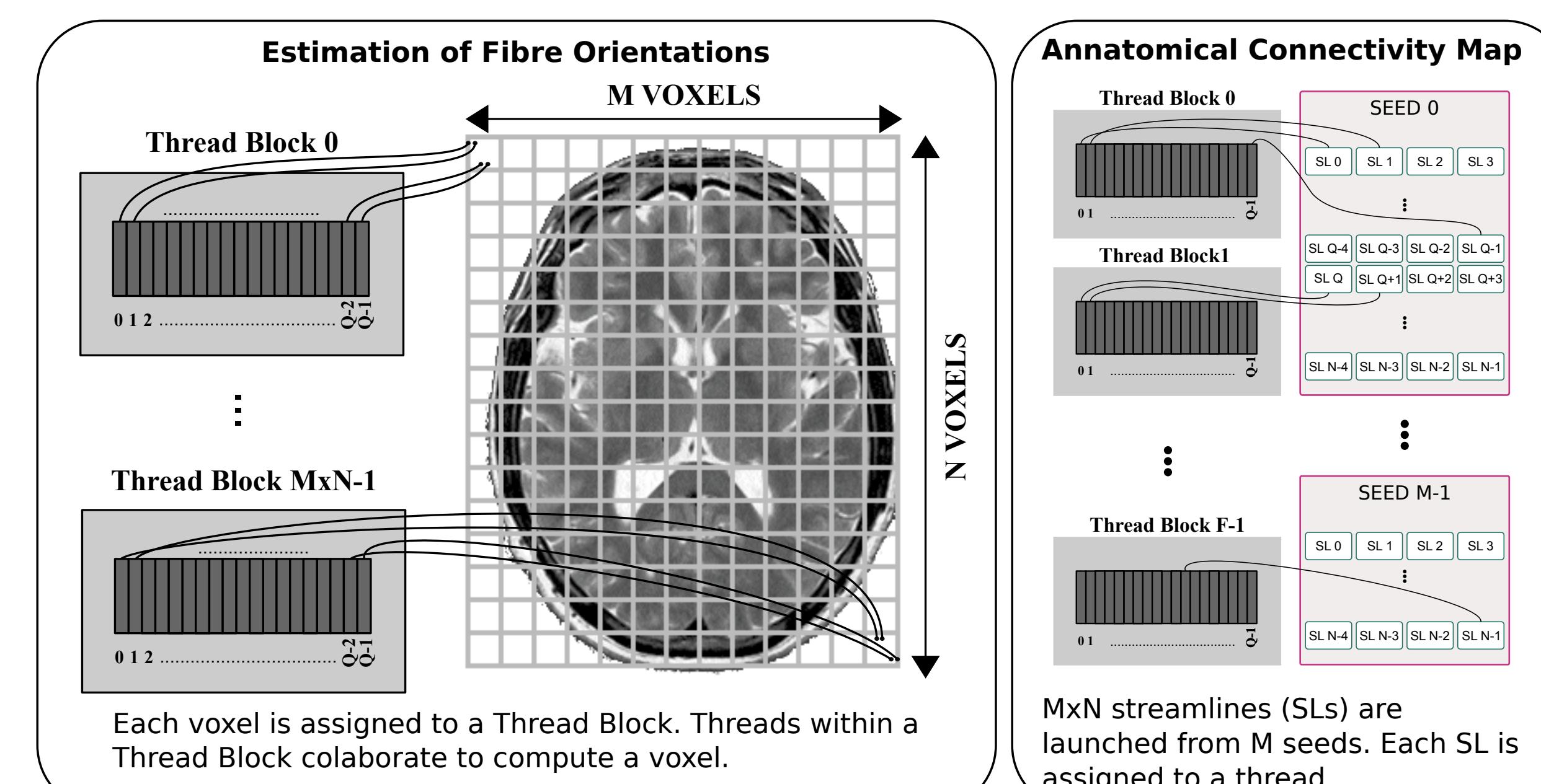
The analysis of dMRI data can require high computational resources:

An inference framework is commonly used to estimate tissue microstructure parameters, such as a distribution of fibre orientations, considering multidimensional data in thousands to millions of voxels.

Subsequently, another framework considers all the estimates as a whole and computes a global anatomical brain connectivity map integrating the brain fibre orientations of continuous voxels. This task is repeated many times in a Monte-Carlo fashion to achieve convergence, launching thousands of streamlines from different seeds.

Overall, the computation times can be significant. For data mining applications that process large dMRI databases, such as data from the HCP (1200 adults) or the UK Biobank Project (100,000 participants), these can limit the potential exploration.

We have developed a novel parallel computational framework for implementing the above tasks.



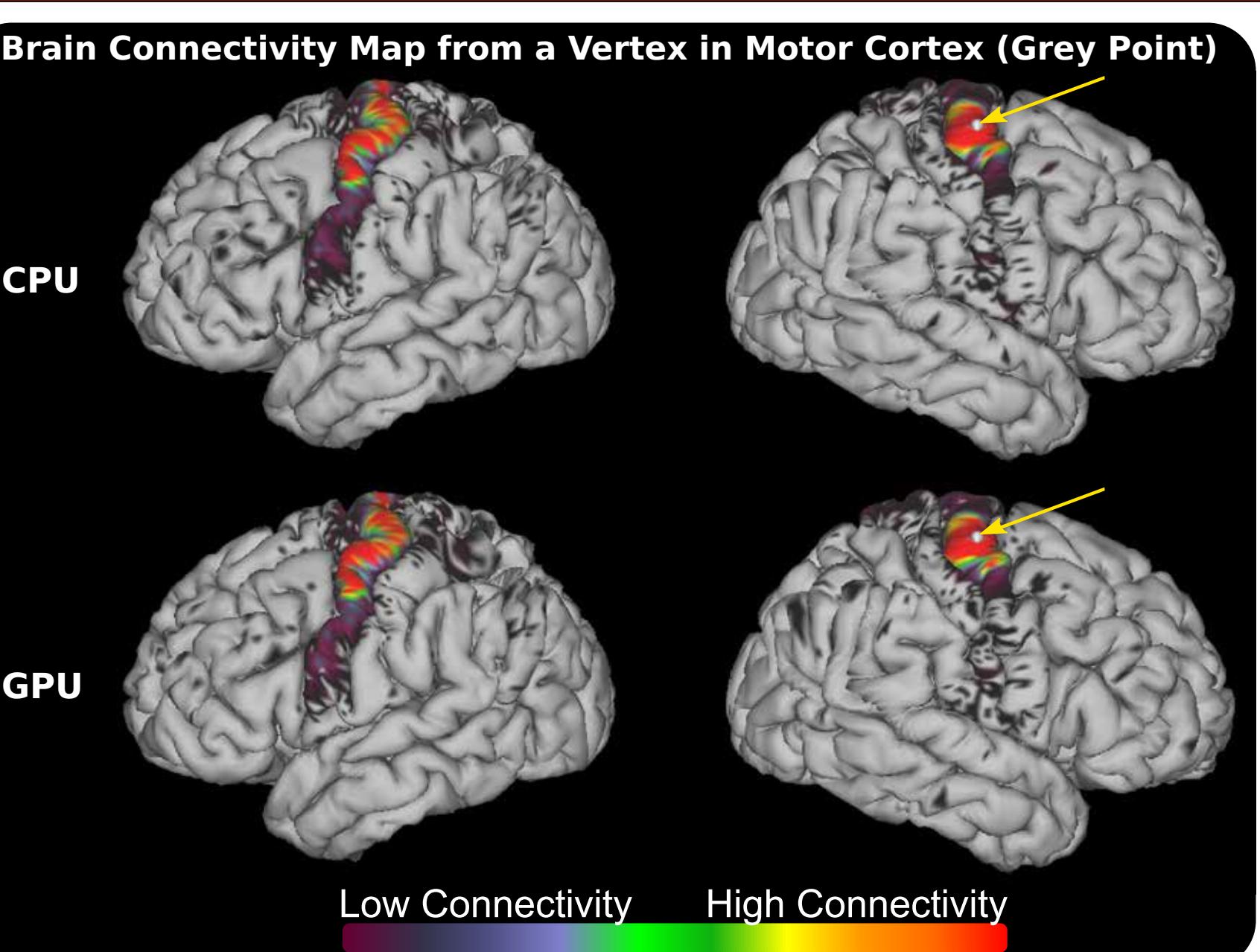
In our design we implement optimisations that lead to remarkable acceleration, including efficient use of GPU memory hierarchy, use of CUDA streams for concurrency, and data structure transformation.

## 3. Results

Only results for the framework that computes the global anatomical brain connectivity map are shown here.

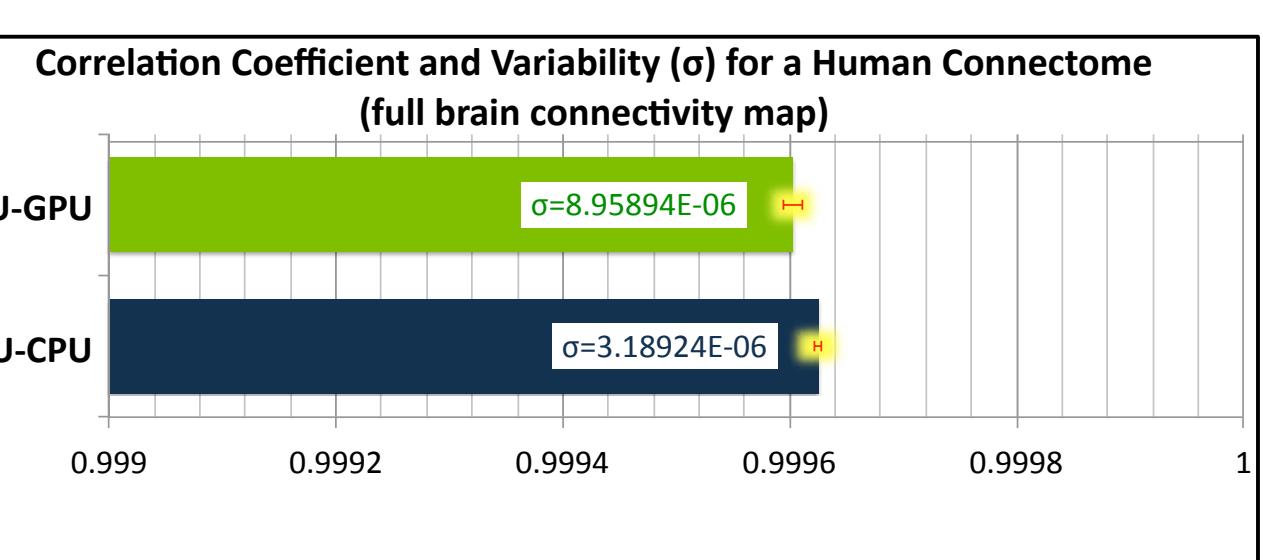
### 3.1 Qualitative Results using a K80 Nvidia GPU

- Anatomical connectivity maps: likelihood of a certain point on the cortical surface being connected to any other cortical point.
- Some major pathways connecting different parts of the brain (Section 1).



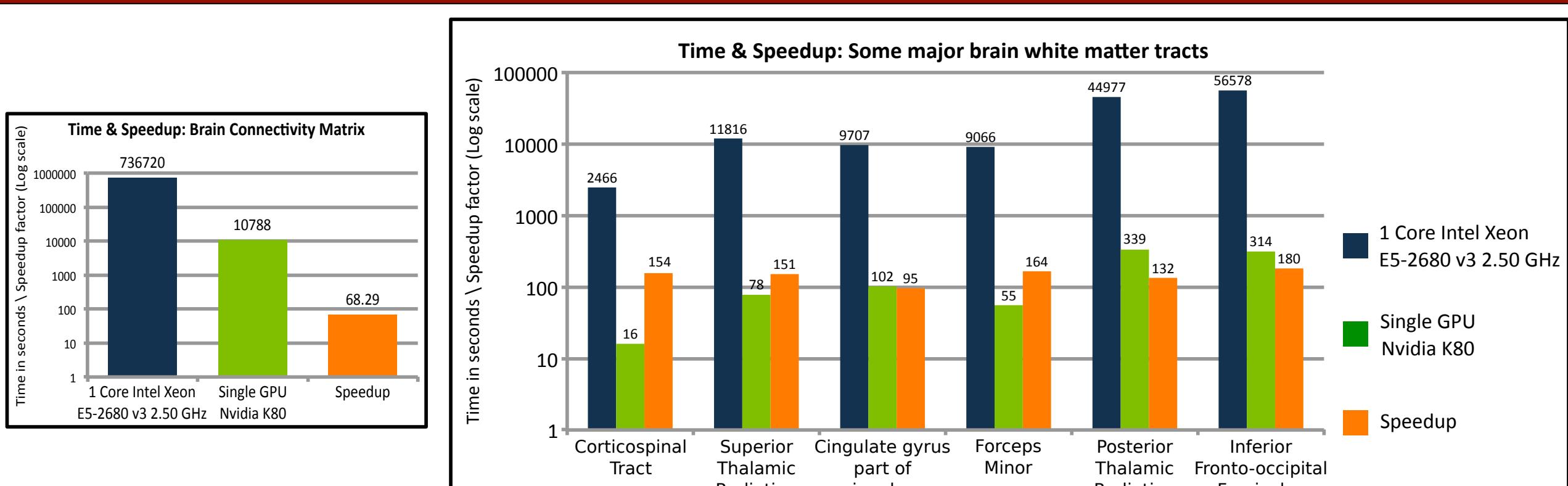
### 3.2. Quantitative Results

Given the stochastic nature of the process we validate our CUDA implementation results by comparing the mean of the correlation coefficients (and their variability) of the results from GPU and several CPU runs, with those from only several CPU runs.



Some major brain white matter tracts	CPU-CPU	CPU-GPU
Left Corticospinal tract	0.9989474	0.9989727
Right Corticospinal tract	0.9995147	0.9995084
Left Superior thalamic radiation	0.9996158	0.9995919
Right Superior thalamic radiation	0.9996538	0.9996358
Left Cingulate gyrus part of cingulum	0.9999928	0.9999930
Right Cingulate gyrus part of cingulum	0.9999873	0.9999889
Forceps minor	0.9997898	0.9997915
Left Posterior thalamic radiation	0.9999715	0.9999701
Right Posterior thalamic radiation	0.9999594	0.9999609
Left Inferior fronto-occipital fasciculus	0.9999660	0.9999662
Right Inferior fronto-occipital fasciculus	0.9999486	0.9999492

### 3.3. Performance Results using a K80 Nvidia GPU



	Brain Connectivity Matrix	White matter tracts	
Global Load Throughput	27.78 GB/s	32.81 GB/s	A few iterations threads of the same warp access to remote memory positions (Uncollapsed): Different brain paths.
Global Load Efficiency	23.11 %	23.56 %	
Global Store Throughput	6.55 GB/s	7.57 GB/s	
Global Store Efficiency	12.58 %	12.62 %	
Instructions executed per cycle (IPC)	1.07	1.15	
Single-precision floating-point operations per second (FLOPS)	12.25 GFLOPS	14.88 GFLOPS	Different brain paths cause divergences.
Single-precision floating-point instructions per second	284.40 GIPS	344.47 GIPS	
Integer instructions per second	389.30 GIPS	477.80 GIPS	
Control instructions per second	51.74 GIPS	61.59 GIPS	A lot of other kind of instructions.