TRX: A community-oriented tractography file format

Francois Rheault[a], Valérie Hayot-Sasson[b], Robert E. Smith[c], Christopher Rorden[d], Jacques-Donald Tournier[e], Eleftherios Garyfallidis[f], Franco Pestilli[g], Fang-Cheng Yeh[h], Chris Markiewicz[i], Matthew Brett[j], Ben Jeurissen[k], Paul A. Taylor[l], Dogu B. Aydogan[m], Derek A. Pisner[n], Serge Koudoro[f], Soichi Hayashi[o], Daniel Haehn[p], Steve Pieper[q], Daniel Bullock[r], Emanuele Olivetti[s], Jean-Christophe Houde[i], Marc Alexandre Coté[i]
Maxime Descoteaux[u,v], Bennett Landman[a,w], Ariel Rokem[x]

Affiliations:

[a]*Electrical and Computer Engineering, Vanderbilt University, Nashville, USA;*
[b]*Department of Computer Science and Software Engineering, Concordia University, Montreal, Canada*
[c]*Florey Institute of Neuroscience and Mental Health, Melbourne, Australia*
[d]*Department of Psychology, University of South Carolina, Columbia, USA*
[e]*Centre for Medical Engineering, King's College London, London, UK*
[f]*Department of Intelligent Systems Engineering, Indiana University, Bloomington, USA*
[g]*Department of Psychology, The University of Texas at Austin, Austin, USA.*
[h]*Department of Neurological Surgery, University of Pittsburgh Medical Center, Pittsburgh, USA*
[i]*Department of Psychology, Stanford University*
[j]*Brain Imaging Center, University of California, Berkeley, Berkeley, CA, USA*
[k]*Department of Physics, University of Antwerp, Antwerp, Belgium*
[l]*Scientific and Statistical Computing Core, National Institute of Mental Health, USA*
[m]*Department of Neuroscience and Biomedical Engineering, Aalto University, Espoo, Finland*
[n]*University of Texas at Austin, Austin, USA*
[o]*Department of Psychological and Brain Sciences, Indiana University, Bloomington, USA*
[p]*Machine Psychology Lab, Department of Computer Science, University of Massachusetts, Boston, USA*
[q]*Department of Neurosurgery, Brigham and Women's Hospital, Harvard Medical School, Boston, United States*
[r]*Center for Magnetic Resonance Research, University of Minnesota*
[s]*Neuroinformatics Laboratory, Fondazione Bruno Kessler, Trento, Italy*
[t]*Microsoft Research Lab, Montréal, Canada*
[u]*Imeka Solutions, Sherbrooke, Canada*
[v]*Sherbrooke Connectivity Imaging Laboratory (SCIL), Département d'Informatique, Université de Sherbrooke, Canada;*
[w]*Vanderbilt University Institute of Imaging, Nashville, USA;*
[x]*Department of Psychology and eScience Institute, University of Washington, Seattle, USA*

Introduction

TRX is a proposed tractography file format designed to bring our community together and facilitate datasets exchange and interoperability. Acting as a next-generation replacement to the **TCK and TRK** [1, 2] as well as less commonly used formats **VTK, FIB, TRAKO, DPY, zFIB, qFIB** [3, 4, 5, 6]. While some of the current file formats are adequate in terms of functionality and performance, each comes with its own set of limitations or peculiarities.

File formats are usually initially developed for their use within their own tool or specific situations. Developers often sought to solve their current challenges, within the bounds of their current computational resources and envision a particular set of ways to use their tool and file format. With tractography becoming more widespread across a variety of subfields and the increasing size and complexity of datasets, limitations are becoming more and more obvious and difficult to work around.

To overcome these limitations and avoid the birth and death of a plethora of file formats in the future, our **community** initiated a discussion to design a new file format and agreed to participate in its **conception**, **development** and if successful: its **adoption**.

Background

Following discussions on Github (https://github.com/nipy/nibabel/issues/942), a videoconference between the developers of several software libraries for tractography, and at the DIPY online group meeting, we identified key features that were desired by the community (both developers and users): It was important not to lose functionalities currently implemented by TCK and TRK (by far the most commonly used file formats); the second most important variable was computational efficiency (both in speed and file size); and third was extensibility so the file format could evolve more easily. Feedback from different groups showed that developers desired features such as being based on open-source, well-maintained libraries and robustness while users desired new functionalities, simplicity and low memory usage.

It was proposed to simply extend an existing file format to reach these goals instead of creating a new one. However, it was determined that it would better convey our community-oriented intentions to work together on the next-generation file formats and agree to participate in its development and endorse it if considered satisfactory.

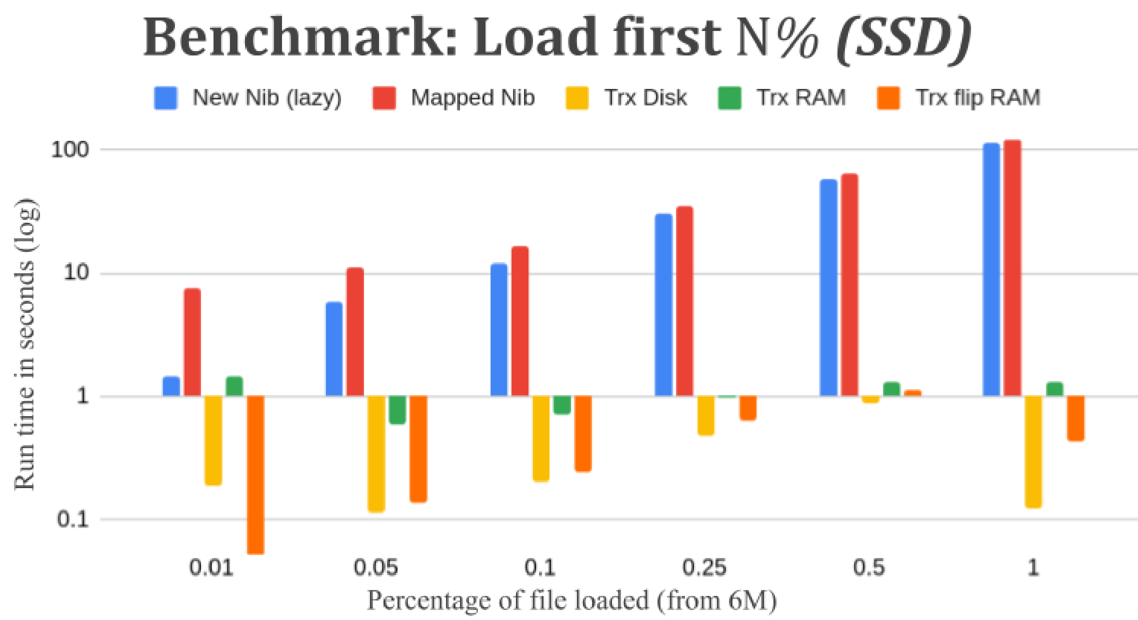
Results

The proposed file format that came to be is **TRX**. Technical specifications can be found in https://github.com/frheault/tractography_file_format/blob/master/trx_file_memmap/specifications.md; a Python implementation (with C++ underway) can be found in https://github.com/frheault/tractography_file_format/tree/master/trx_file_memmap; and various discussions related to the design can be found in https://github.com/frheault/tractography_file_format/issues. The file format favors simplicity and human-readability when it comes to its architecture, the data is organized following a naming convention that conveys dimensionality and data type. TRX allows storage of data per streamline, data per vertex (functionalities from TRK) and sparse indices (groups) as well as data per group. We reached an agreement to enforce RASMM (world/scanner coordinates) for tractography's coordinate system for streamlines (like TCK). The file format supports random-access with minimal upfront RAM requirement, and other low-level operations with virtually no RAM requirements, by using a data & offsets approach (similar to VTK9) and raw binary arrays that allows memory mapping. The file format supports linearisation (a form of streamline compression), controls over all data types and, if needed, ZIP compression (lossy or lossless)

Benchmarking of simple operations such as loading partial file (Figure 1), concatenation, applying an affine transformation, saving (Figure 2) was performed using the Python implementation and compared to Nibabel. Due to the implementation using memmap, the operations were tested from a file on disk (SSD) to RAM for loading and from RAM to disk (SSD) for saving. A C++ implementation of the TRX is planned for benchmarking against a C++ TCK reader/writer.

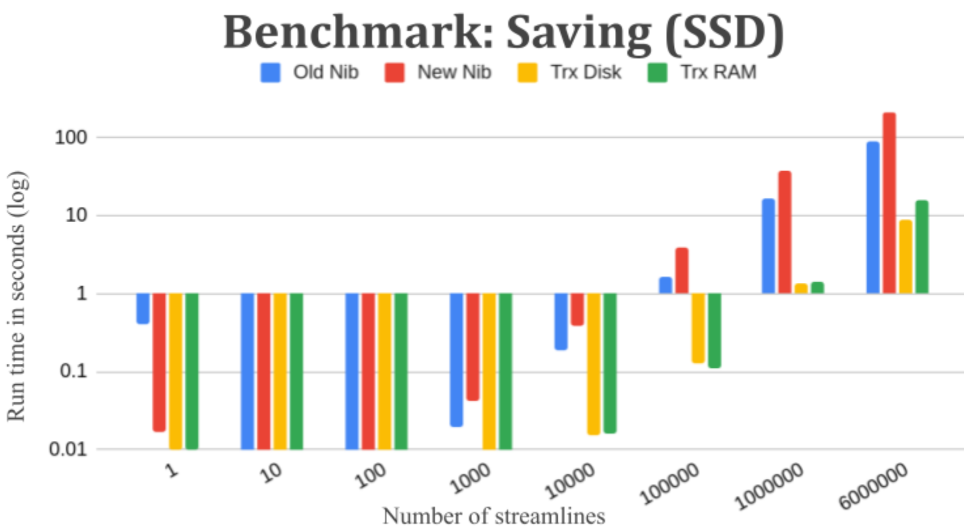
Conclusion

TRX is a community-based proposal for the next-generation file format. The goals are to retain the best functionalities of TCK and TRK, add new features useful to the community, be better adapted to current computational needs and be easily extendable in order to follow the evolution of our field.



Percentage	**1%**	**5%**	**10%**	**25%**	**50%**	**100%**
New Nib (lazy)	1.4395	5.83541	12.16036	30.59931	57.86782	115.64275
Mapped Nib	7.62627	11.18746	16.69855	34.7145	63.28102	120.81939
Trx Disk	0.18648	0.1155	0.20596	0.47349	0.86576	0.12334
Trx RAM	1.47415	0.59769	0.70536	0.98844	1.2861	1.31336
Trx flip RAM	0.04962	0.13847	0.24648	0.62824	1.11471	0.43175

Figure 1: Load times (in seconds) for the first N% streamlines of a large file (6M streamlines). To facilitate visualization and comparison, the graph is shown on a logarithmic scale. *New Nib* is the Nibabel API (version 3.0, TRK), Mapped Nib is an implementation that allows random access (available on https://github.com/emanuele/load_trk, TRK). *TRX Disk* is in fact simply the memory mapping initialization, there is no "loading". *TRX RAM* is the memory mapping initialization, transfer of all the data from disk to RAM followed by the subsampling (heavy memory usage) and *TRX flip RAM* is the memory mapping, transfer of only the first N% to RAM (low memory usage). Nibabel API being a Python implementation it is likely slower than the implementation of TrackVis [2] or MI-Brain [7], but this showcases the advantage of memory mapping for slower scripting languages.



# Streamlines	1	10	100	1000	10000	100000	1000000	6000000
Old Nib	0.40020	0.00032	0.00190	0.01907	0.18864	1.60935	16.31758	89.18324
New Nib	0.01658	0.00126	0.00481	0.04265	0.37821	3.85463	38.15885	210.43288
Trx Disk	0.01014	0.00373	0.00385	0.00544	0.01536	0.12890	1.33239	9.03163
Trx RAM	0.00992	0.00371	0.00385	0.00524	0.01588	0.11144	1.43796	15.53585

Figure 2: Saving times (in seconds) of various file sizes from 1 to 6M streamlines. To facilitate visualization and comparison, the graph is shown on a logarithmic scale. Old Nib is an older python implementation available in DIPY [8] that does not support data per streamlines/points (deprecated, TRK). New Nib is the Nibabel API (version 3.0, TRK). Due to the usage of memory mapping, TRX Disk is simply a copy-paste operation to show the baseline disk speed. TRX RAM is the saving operation from all data in RAM to disk (including memmaps initialization and writing).

References

[1] Tournier, J-Donald, Robert Smith, David Raffelt, Rami Tabbara, Thijs Dhollander, Maximilian Pietsch, Daan Christiaens, Ben Jeurissen, Chun-Hung Yeh, and Alan Connelly. "MRtrix3: A fast, flexible and open software framework for medical image processing and visualisation." *Neuroimage* 202 (2019): 116137.

[2] Wang, Ruopeng, Thomas Benner, Alma Gregory Sorensen, and Van Jay Wedeen. "Diffusion toolkit: a software package for diffusion imaging data processing and tractography." In *Proc Intl Soc Mag Reson Med*, vol. 15, no. 3720. 2007.

[3] Yeh, Fang-Cheng, Timothy D. Verstynen, Yibao Wang, Juan C. Fernández-Miranda, and Wen-Yih Isaac Tseng. "Deterministic diffusion fiber tracking improved by quantitative anisotropy." *PloS one* 8, no. 11 (2013): e80713.

[4] Mercier, Corentin, Sylvain Rousseau, Pietro Gori, Isabelle Bloch, and Tamy Boubekeur. "QFib: fast and efficient brain tractogram compression." *Neuroinformatics* 18 (2020): 627-640.

[5] Haehn, Daniel, Loraine Franke, Fan Zhang, Suheyla Cetin-Karayumak, Steve Pieper, Lauren J. O'Donnell, and Yogesh Rathi. "Trako: Efficient transmission of tractography data for visualization." In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 322-332. Springer, Cham, 2020.

[6] Rheault, Francois, Jean-Christophe Houde, and Maxime Descoteaux. "Visualization, interaction and tractometry: dealing with millions of streamlines from diffusion MRI tractography." *Frontiers in neuroinformatics* 11 (2017): 42.

[7] Rheault, Francois, Jean-Christophe Houde, Nil Goyette, F. Morency, and Maxime Descoteaux. "MI-Brain, a software to handle tractograms and perform interactive virtual dissection." In *Proceedings of the ISMRM Diffusion study group workshop, Lisbon*. 2016.

[8] Garyfallidis, Eleftherios, Matthew Brett, Bagrat Amirbekian, Ariel Rokem, Stefan Van Der Walt, Maxime Descoteaux, and Ian Nimmo-Smith. "Dipy, a library for the analysis of diffusion MRI data." *Frontiers in neuroinformatics* 8 (2014): 8.