

Think Global, Act Local; Projectome Estimation with BlueMatter^{*}

Anthony J. Sherbondy¹, Robert F. Dougherty²,
Rajagopal Ananthanarayanan¹, Dharmendra S. Modha¹,
and Brian A. Wandell²

¹ IBM Almaden Reserach Center, Almaden, USA
`anthony.sherbondy@stanford.edu`

² Psychology Department, Stanford University, USA

Abstract. Estimating the complete set of white matter fascicles (the projectome) from diffusion data requires evaluating an enormous number of potential pathways; consequently, most algorithms use computationally efficient greedy methods to search for pathways. The limitation of this approach is that critical global parameters - such as data prediction error and white matter volume conservation - are not taken into account. We describe BlueMatter, a parallel algorithm for global projectome evaluation, which uniquely accounts for global prediction error and volume conservation. Leveraging the BlueGene/L supercomputing architecture, BlueMatter explores a massive database of 180 billion candidate fascicles. The candidates are derived from several sources, including atlases and mutliple tractography algorithms. Using BlueMatter we created the highest resolution, volume-conserved projectome of the human brain.

1 Introduction

The white matter of the human brain comprises more than $150km$ of long-range myelinated connections [1]. Understanding the architecture of these long-range projections (the projectome) is important for understanding brain function [2]. Diffusion weighted imaging fiber tractography (DWI-FT) is the only non-invasive method for studying the human brain projectome.

While there has been great progress in developing fiber tracking techniques [3,4,5], there is wide agreement that current methods fail in many specific cases [6,7,4]. A limitation is that current algorithms find pathways using greedy techniques; that is, the algorithms make decisions based on individual tracts without considering the entire projectome. Further, current algorithms do not optimize

^{*} Sponsored by Defense Advanced Research Projects Agency, Defense Sciences Office (DSO), Program: Systems of Neuromorphic Adaptive Plastic Scalable Electronics (SyNAPSE), Issued by DARPA/CMO under Contract No. HR0011-09-C-0002. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressly or implied, of the Defense Advanced Research Projects Agency or the U.S. Government. Also supported by NEI EY01500.

the projectome to fit the original diffusion data. Finally, they ignore important physical constraints, such as the volume occupied by the estimated fascicles.

We introduce the BlueMatter algorithm to address these limitations. BlueMatter takes as input fascicles derived from multiple tractography algorithms. It searches for an optimal combination of these fascicles (the projectome) subject to two error terms. The first term compares the predicted and measured diffusion-weighted images. We refer to this constraint as diffusion-fitting (see also [8,9]). The second term accounts for an upper limit on the fascicle volume. We refer to this constraint as volume regularization; this term is an important physical constraint that helps resolve the ill-posed inverse problem of finding a projectome that minimizes diffusion-fitting error.

Accounting for diffusion-fitting and fiber count was discussed by Zhang and Laidlaw [10]; however, their technique did not address fascicle volume estimation and was limited to fascicles derived from deterministic algorithms.

BlueMatter is the first algorithm to produce a human brain projectome that combines diffusion-fitting and volume regularization. BlueMatter combines fascicle estimates from deterministic and stochastic DWI-FT algorithms; uniquely integrating algorithms with different strengths. Because BlueMatter has a very large and complex search space, the algorithm is highly parallelized and takes advantage of modern distributed computing architectures. Using this algorithm on a 2048-processor BlueGene/L supercomputer with 0.5 TB of memory we create the highest resolution, physically plausible projectome to date.

Table 1. Symbols and terms

Name	Description (Values)
C	Candidate fibers collected from many tractography algorithms
P, P_i	Current projectome estimate and portion of estimate intersecting i th voxel
E	Total projectome estimate error
E_1	Error between diffusion data prediction and measurement
E_2	Error from over filling voxels with white matter estimates
λ	Balance between E_1 and E_2 to compute E (0.2)
$A_{k,i}$	Diffusion attenuation measurement along the k th axis at the i th voxel
$\hat{A}_{k,i}$	Predicted diffusion attenuation along the k th axis at the i th voxel
$\nu_{C,i}, \nu_{U,i}$	Estimated volume of a voxel attributed to CSF and unoriented tissue
$\nu_{f,i}$	Estimated volume of a voxel attributed to the fascicle tissue
ν_T	Target volume of white matter
d_C, d_U	Diffusivity in all directions within canonical CSF and unoriented tissue compartments (3.1, 0.85 $\mu m^2/ms$)
d_a, d_r	Diffusivity along the axial axis and radial axes within a canonical white matter fascicle compartment (2.0, 0.275 $\mu m^2/ms$)
D_f	Canonical white matter diffusion tensor with eigenvalues d_a, d_r and d_r , and first eigenvector oriented with fascicle tangent
σ	Standard deviation of the data estimated across all brain voxels
τ_L	Convergence threshold for local improvements (100 in 5000 iterations)
τ_G	Convergence threshold for global improvements (50 in 50 iterations)
Δ_N	Amount of time to find their next neighbor state (5 minutes)

2 Algorithm

BlueMatter searches a massive collection of fascicle candidates C to select a projectome P . The quality of a projectome is evaluated by a global error function E . C is intended to be very large and contain a superset of the real fascicles. The collection C must be refined because it (a) contains many implausible fascicles, (b) reflects the biases of the generating tractography algorithms, and (c) is not optimized to fit the data or satisfy volumetric constraints. BlueMatter leverages the BlueGene/L architecture to search for an optimal projectome estimate from the enormous space of possible projectomes within C .

Table 1 defines key symbols and terms.

Error Terms. The BlueMatter algorithm penalizes a projectome, P , with a global error metric, E ,

$$E = (1 - \lambda)E_1 + \lambda E_2. \tag{1}$$

The error E is a convex combination of E_1 , which measures the difference between the predicted and observed diffusion weighted images, and E_2 , the amount the volume is overfilled. The parameter λ modulates the balance between E_1 and E_2 and is selected empirically (Section 3). We seek solutions that predict the data to within the measurement noise; we also seek solutions with no more fascicles than a given voxel volume allows.

The projectome estimates the amount of volume in a voxel occupied by fascicles and the remaining space is then filled with isotropic diffusion according to the ratio $\nu_{U,i}/\nu_{C,i}$. The diffusion sensitization strength b is set by the MRI scanning sequence. The MRI sequence also specifies the axes of diffusion sensitivity or q_k , where $k \in [1, K]$.

We define the diffusion estimation error, E_1 , and the local volume overfilling error, E_2 , at voxel i as

$$E_{1,i} = \frac{1}{K} \sum_{k=1}^K \frac{(\hat{A}_{k,i} - A_{k,i})^2}{\sigma^2},$$

$$E_{2,i} = \begin{cases} 0 & \text{if } \sum_{f \in P_i} \nu_{f,i} < \nu_T \\ \sum_{f \in P_i} \nu_{f,i} - \nu_T & \text{otherwise.} \end{cases} \tag{2}$$

Both E_1 and E_2 are the sums across voxels of $E_{1,i}$ and $E_{2,i}$, respectively. E_1 is normalized by the variance of the noise and thus represents error in units of data variance (or transformed into standard deviations). E_2 is in units of mm^3 .

We compute E_1 based on a model for the predicted diffusion weighted image attenuation at voxel i and direction k , $\hat{A}_{k,i}$. BlueMatter uses the powder average of the separate compartments [11]. This term can be calculated from the projectome fibers passing through this voxel, P_i

$$\hat{A}_{k,i} = \frac{A_{0,i}}{\nu_{C,i} + \nu_{U,i} + \sum_{f \in P_i} \nu_{f,i}} (\nu_{C,i} \exp(-bd_C) +$$

$$\nu_{U,i} \exp(-bd_U) + \sum_{f \in P_i} \nu_{f,i} \exp(-bq_k^T D_{f,i} q_k). \quad (3)$$

Parallel Candidate Generation. The set of candidate fibers is created by combining estimates from multiple tractography algorithms. It is intended that C contain a superset of fascicles; hence, we use a large candidate set derived from STT [3], TEND [12] and ConTrack [7] algorithms.

Projectome Search Overview. BlueMatter searches through C using a parallel stochastic hill climbing algorithm with multiple-restarts. A standard steepest ascent hill climbing algorithm would take the current state and every one of its neighbors; and choose the next state to be the neighbor that increases the inverse of the cost function the most [13]. By contrast, stochastic hill climbing algorithm only looks at a subset of the current state's neighbors and selects the next state that provides the largest improvement. For large neighborhoods and high data dimensionality, stochastic hill climbing substantially reduces the search time or in this case makes search feasible.

To reduce the likelihood of being stuck in local maxima, ridges or plateaus, groups of BlueGene/L processors are devoted to independent stochastic hill climbing processes (multiple-restarts) and the best resulting projectome is selected after each group converges. This follows the rule of thumb that when the terrain of the cost function is unknown it is beneficial to devote resources to covering more variable states rather than designing complex local movements. Thus, BlueMatter simultaneously exploits inherent data parallelism in the stochastic hill climbing algorithm and process parallelism in the multiple-restart approach.

Stochastic Search Implementation. For each restart, a subset of processors is selected to search the current projectome neighborhood. Each processor within a subset is assigned a random samples of C . The processors perform the following sequential optimization procedure for a specified amount of processor time.

To start, each processor selects a random voxel group. The processor then alternates between fiber addition and subtraction operations using the selected group. During addition, fibers that intersect the current voxel group are sampled from C . If the addition reduces the total error, the fiber is retained. During subtraction, fibers that intersect the voxel group are randomly sampled from P . If removing the fiber reduces the total error, the fiber is deleted. The algorithm alternates between addition and subtraction for a voxel group until the error reduction slows below a predefined threshold τ_L . After local improvement slows, a new voxel group is selected and the local optimization begins again.

Each processor optimizes its subset of candidate pathways, and processors within the subset exchange their current projectome estimate. After a small amount of processor time Δ_N has passed each processor points to a unique guess for the next projectome. The projectome with the lowest error is selected as the next state.

The volume constraint limits the size of the projectome state information, which is exchanged between processors quickly. Efficient communication is crucial to achieving the benefit of the massive and distributed database.

Multiple-Restart. The multiple-restart is achieved by devoting multiple sets of processors to independent runs of the stochastic hill climbing algorithm. When the rate of improvement of projectome error of each of the separate stochastic hill climbing groups slows below a threshold τ_G , the groups have converged to a projectome estimate. The projectome with the least error is selected as the optimal projectome.

The Initial Projectome. BlueMatter's initial projectome estimate is a subset of the STT and TEND fibers in the Mori human white matter atlas [14]; these are selected to minimize the error terms (see 2). This is a very small subset of candidate fibers (10,000) that are repeatedly found with previous algorithms.

3 Results

We first used a synthetic data set to demonstrate that BlueMatter recovers ground truth orientation and volume from diffusion-weighted images. We then used BlueMatter to estimate a projectome in human DWI data.

Synthetic. The synthetic dataset (Figure 1A) consists of two fascicle bundles. One bundle (blue) fills 60% of each voxel with fibers coherently organized in one direction. A second bundle (green) occupies 40% of each voxel with fascicles oriented perpendicular to the blue bundle. At the intersection, voxels are 100% filled with white matter from the two bundles. After accounting for the white matter, unfilled space within the bundles is filled with isotropic diffusion of the same mean diffusivity as white matter. The ends of both bundles are capped with voxels containing isotropic diffusion at the mean diffusivity of gray matter.

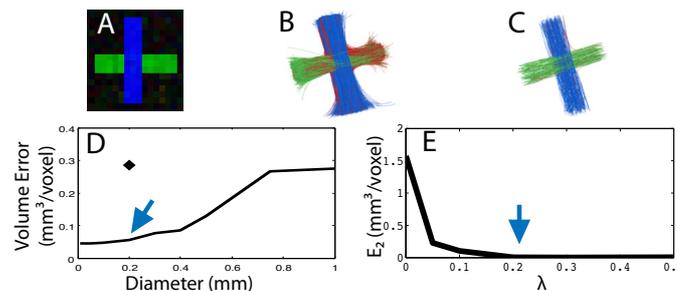


Fig. 1. Projectome estimation from synthetic data. (A) A synthetic data set comprising two fiber groups (green and blue). (B) A 3D view of the fascicle candidate database, C . (C) The BlueMatter projectome estimate P accurately recovers the synthetic fascicle ratio. (D) Small fascicle diameters produce a projectome with more accurate volume estimates. If probabilistic candidate pathways are excluded from C , the error is significantly worse (black diamond). (E) Increasing λ has a negligible effect on the diffusion-fitting error, E_1 (not shown) and significantly reduces E_2 . Blue arrows (D and E) indicate the BlueMatter parameters used in the experiments.

The other voxels in the 20x20x20 volume are modeled as CSF. Diffusion data were simulated using the partial compartment model (Eq. 1) measured in six directions (plus a $b=0$ measurement) with 8 repeats. Rician noise with the same standard deviation as the data used in Section 3.2 ($\sigma=48$) was added to the magnitude attenuation signal.

Figure 1B shows the candidate set C derived using the methods described in Section 2. The fascicles are colored according to which ends they reach with blue and green matching the original; red fascicles are wrong turns. The candidate fascicles do not have the correct volume ratios and include incorrect fascicles. After running BlueMatter on this small synthetic data set, the proper blue-green ratio is returned and only a small number of wrong turns remain (Figure 1C).

We used this synthetic data to set (a) the fascicle diameter and (b) the balance between the two error terms. We evaluated the accuracy of the volume match by varying the modeled fascicle diameter (Figure 1D). A small fascicle diameter, on the order of $200\mu\text{m}$ (blue arrow), produces a good match to the synthetic projectome volume. Including probabilistic candidate pathways is important to achieving an accurate match. Without these candidate fascicles the volume estimate is significantly worse (black diamond).

Altering the balance between the diffusion-fitting and volume error terms has a negligible effect on the diffusion-fitting error (not shown). Setting the balance parameter, λ , reduces the volume error (E_2) error. We selected $\lambda = 0.2$ (blue arrow) for the experiments. With these parameters, BlueMatter recovers the volume and orientation of the synthetic white matter.

Human Projectome. In the second experiment we used BlueMatter to estimate a human projectome. First, 44,000 cortical and subcortical gray matter voxels were semi-automatically identified using FAST [15] and FIRST [16]. A set C containing 180 billion candidate fascicles was constructed. The set included fascicles from each and every gray matter voxel (Section 2). The BlueMatter search algorithm converged on a projectome with 200,000 fascicles after 9 days

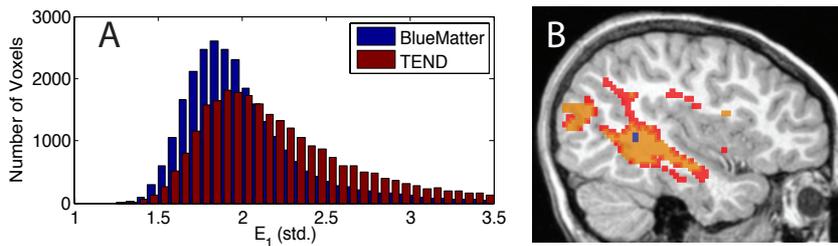


Fig. 2. Comparison between the BlueMatter and TEND human projectomes. (A) Diffusion-fit errors E_1 for BlueMatter (blue) and TEND (red) projectomes. BlueMatter reduces diffusion-fit error in most voxels. (B) A mid-temporal sagittal section of the human brain; the color overlay shows voxels overfilled by at least 2x (red) and 3x (orange) with the TEND projectome. The BlueMatter projectome only overfills two voxels (blue) by 2x and none by 3x.

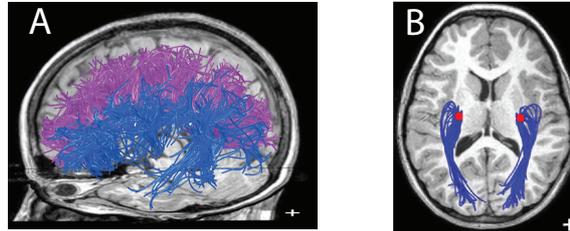


Fig. 3. Important fiber tracts within the BlueMatter projectome. (A) Pathways project from corpus callosum to lateral surface (blue). (B) Optic radiation (blue) connecting lateral geniculate nucleus (red) to primary visual cortex.

of compute time on the 2048-processor BlueGene/L supercomputer. We also computed a projectome using the TEND algorithm (see Izhikevich et al. [17]). The TEND projectome comprises 300,000 fascicles.

To estimate the white matter volume in each voxel of the TEND projectomes, we searched over fascicle model diameters ($20\text{-}300\mu\text{m}$) to find the diameter that has the lowest diffusion-fit error within the white matter core (linearity > 0.3). The optimal TEND diameter is the same as BlueMatter's diameter ($200\mu\text{m}$).

Using the optimal fascicle diameter, we could estimate the diffusion-fitting (E_1) and volume conservation (E_2) errors for both algorithms. The mean E_1 errors are 2.4 (mode 2.1) and 2 (mode 1.8) for the TEND and BlueMatter projectomes, respectively (Figure 2A). Using a scatter plot (not shown), we verified that BlueMatter E_1 error is better at nearly every voxel. The smallest possible mean E_1 error can be found by fitting a tensor to the data at each voxel: This lower bound is 1.7 (mode 1.6), which is slightly better than the BlueMatter fit.

We assessed the volume conservation error by counting the number of overfilled voxels (Figure 2B). The BlueMatter projectome has only 12 (200) voxels overfilled by more than 3x (2x); the TEND projectome has 3500 (8200) overfilled voxels. We explored compensating for TEND's overfilling by reducing the fascicle diameter. Reducing the diameter to $100\mu\text{m}$ reduces the overfilling error to match BlueMatter, but the mean E_1 value rises to 2.7 and many core white matter voxels, such as the corona radiata, are emptied of white matter.

Important Fiber Tracts. The BlueMatter human projectome contains the major intra-hemispheric white matter pathways as reported by Wakana et al. [14]. The projectome also contains fasciculi not present in the atlas a) fascicles that connect the corpus callosum to lateral cortex (Figure 3A, blue) and b) optic radiation (Figure 3B). These tracts are rarely found using local, greedy methods.

4 Conclusion

The BlueMatter algorithm searches across a pool of 180 billion candidate fascicles, drawn from many sources, to find a projectome. The highly parallelized implementation takes advantage of the 2048-processor BlueGene/L supercomputer.

BlueMatter is the first algorithm that simultaneously fits two global parameters of the projectome: the diffusion data (E_1) and volume conservation (E_2). The algorithm successfully identifies the optic radiation and other pathways that are frequently missed by local, greedy methods.

References

1. Braitenberg, V., Schüz, A.: *Cortex: Statistics and Geometry of Neuronal Connectivity*. Springer, Berlin (1998)
2. Kasthuri, N., Lichtman, J.W.: The rise of the 'projectome'. *Nat. Methods* 4(4), 307–308 (2007)
3. Basser, P.J., Pajevic, S., Pierpaoli, C., Duda, J., Aldroubi, A.: In vivo fiber tractography using DT-MRI data. *Magn. Reson. Med.* 44(4), 625–632 (2000)
4. Mori, S., van Zijl, P.C.: Fiber tracking: principles and strategies - a technical review. *NMR Biomed.* 15(7-8), 468–480 (2002)
5. Parker, G.J., Alexander, D.C.: Probabilistic anatomical connectivity derived from the microscopic persistent angular structure of cerebral tissue. *Philos Trans. R. Soc. Lond. B. Biol. Sci.* 360(1457), 893–902 (2005)
6. Miller, N.R.: Diffusion tensor imaging of the visual sensory pathway: are we there yet? *Am. J. Ophthalmol.* 140(5), 896–897 (2005)
7. Sherbondy, A.J., Dougherty, R.F., Ben-Shachar, M., Napel, S., Wandell, B.A.: ConTrack: Finding the most likely pathways between brain regions using diffusion tractography. *Journal of Vision* 8(9), 1–16 (2008)
8. Poupon, C., Clark, C.A., Frouin, V., Regis, J., Bloch, I., Le Bihan, D., Mangin, J.: Regularization of diffusion-based direction maps for the tracking of brain white matter fascicles. *Neuroimage* 12(2), 184–195 (2000)
9. Peled, S., Friman, O., Jolesz, F., Westin, C.F.: Geometrically constrained two-tensor model for crossing tracts in DWI. *Magn Reson Imaging* 24(9), 1263–1270 (2006)
10. Zhang, S., Laidlaw, D.H.: Sampling DTI fibers in the human brain based on DWI forward modeling. In: *Conf. Proc. IEEE Eng. Med. Biol. Soc.*, vol. 1, pp. 4885–4888 (2006)
11. Basser, P.J., Jones, D.K.: Diffusion-tensor MRI: theory, experimental design and data analysis - a technical review. *NMR Biomed.* 15(7-8), 456–467 (2002)
12. Lazar, M., Weinstein, D.M., Tsuruda, J.S., Hasan, K.M., Arfanakis, K., Meyerand, M.E., Badie, B., Rowley, H.A., Haughton, V., Field, A., Alexander, A.L.: White matter tractography using diffusion tensor deflection. *Hum. Brain Mapp.* 18(4), 306–321 (2003)
13. Russell, S., Norvig, P.: *Artificial Intelligence: A Modern Approach*, 2nd edn. Prentice-Hall, Englewood Cliffs (2003)
14. Wakana, S., Jiang, H., Nagae-Poetscher, L.M., van Zijl, P.C., Mori, S.: Fiber tract-based atlas of human white matter anatomy. *Radiology* 230(1), 77–87 (2004)
15. Zhang, Y., Brady, M., Smith, S.: Segmentation of brain MR images through a hidden markov random field model and the expectation-maximization algorithm. *IEEE Trans. Med. Imaging* 20(1), 45–57 (2001)
16. Patenaude, B., Smith, S., Kennedy, M.: Improved surface models for FIRST. In: *Proc. Hum. Brain Mapp.* (2008)
17. Izhikevich, E.M., Edelman, G.M.: Large-scale model of mammalian thalamocortical systems. *Proc. Natl. Acad. Sci. USA* 105(9), 3593–3598 (2008)