Toward the Human Nanoscale Connectome: Neuronal Morphology Format, Modeling, and Storage Requirement Estimation

Wieslaw L. Nowinski[1]

[1] Sano Centre for Computational Personalised Medicine, Krakow, Poland
w.nowinski@sanoscience.org
www.wieslawnowinski.com

Abstract. The human brain is an enormous scientific challenge. Knowledge of the complete map of neural connections (connectome) is essential for understanding how neural circuits encode information and the brain works in health and disease. Nanoscale connectomes are created for a few small animals but not for human. Moreover, existing models and data formats for neuron morphology description are "merely" at the microscale.

This work (1) formulates a complete set of morphologic parameters of the entire neuron at the nanoscale and introduces a new neuronal nanoscale data format; (2) proposes four geometric neuronal models: straight wireframe, enhanced wireframe, straight polygonal, and enhanced polygonal, based on the introduced neuronal format; and (3) estimates storage required for these neuronal models.

The straight wireframe model requires 18 petabytes (PB). The parabolic wireframe model needs 36 PB and the cubic model 54 PB. The straight polygonal model requires 24 PB. The parabolic polygonal model needs 48 PB and the cubic model 72 PB.

To my best knowledge, this is the first work providing for the human brain (1) the complete set of neuronal morphology parameters, (2) neuronal nanoscale data format, and (3) storage requirement estimation for volumetric and geometric neuronal morphology models at the micro and nanoscales. This work opens an avenue in human brain nanoscale modeling enabling the estimation of computing resources required for the calculation of the nanoscale connectome.

Keywords: Human Brain, Modeling, Storage, Big Data, Neuron, Dendrite, Axon, Nanoscale, Connectome.

1 Introduction

The human brain empowers each of us with enormous functionality enabling perception, locomotion, behavior, emotion, and higher functions, such as cognition, attention, motivation, learning, language, and memory. It performs automatic functions for monitoring, controlling, and maintaining our whole body as well as ensuring survival. The brain determines our personality and is the source of our creativity in problem solving, construction of tools, doing research, development of technology, and artistic creation.

ICCS Camera Ready Version 2023
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-36021-3_66

Understanding the brain structure, function, and dysfunction is an enormous scientific challenge, a critical social need, and a great market opportunity. Though the average human brain has a volume of only 1,400 cm^3 it is the most complex living organ in the known universe with approximately a hundred billion neurons and a thousand trillion connections. Moreover, societies are aging. One-third of the world's adult population suffers from neurological diseases. Brain diseases are the most common and account for 13% of all diseases. The cost of neurological diseases is huge and increasing. Hence, brain research is the next huge technological wave after the space conquest and numerous large-scale projects are underway to uncover the brain's mysteries [1].

One of the key challenges is to generate a connectome at various spatial scales, a complete map of neural connections (circuits) in the brain. This knowledge is essential for understanding how neural circuits encode information and, consequently, how the brain works in health and disease [2]. This will also enable the development of applications in non-medical areas such as neuromorphic computing, artificial intelligence, intelligent machines, and energy-efficient computation. Nanoscale connectomes have been completed only for very small brains including nematode *Caenorhabditis elegans* [3], larva *Ciona intestinalis* [4], and *Drosophila* [5]. At the macroscale, the human connectome has been developed by providing anatomical and functional connectivity [6]. Anatomical connectivity employs diffusion magnetic resonance imaging techniques including diffusion tensor and diffusion spectrum imaging whereas functional connectivity exploits functional magnetic resonance imaging. However, the human nanoscale connectome has not yet been created and here I address the feasibility of producing one.

One of the most critical obstacles in obtaining complete human brain maps at micro and nanoscales is a prohibitively long overall acquisition time. For instance, using the same brain imaging protocol as was employed for *Drosophila* [5] would take an estimated 17 million years to image the whole human brain [7]. Another coarse estimate assesses that the complete reconstruction of a single 1 mm^3 of the cortex would take 10,000 man-years [8] and cost above $100 million [2].

Owing to the progress in imaging, the acquisition time substantially decreases. In particular, by employing synchrotron X-ray tomography the whole human brain acquisition time at the sub-cellular level is estimated to be reduced to a few years [9]. This promising modality, similar to standard computed tomography (CT), provides tissue imaging but with a short wavelength and much higher spatial resolution up to a single synapse [10]. For instance, by employing synchrotron tomography, whole brain imaging was performed for about 16 hours at a 25 μm spatial resolution with 16-bit voxels [11]. This amount of spatial resolution enables imaging of neuronal cell bodies, however, it is insufficient to demonstrate complete neurons with their dendrites and axons. For example, for the human globus pallidus, the average cell body diameter is 33 μm, the proximal dendrite diameter is 4 μm, and the distal dendrite diameter is between 0.3-0.5 μm. Generally, the average diameter of the axon is 1 μm and the synapses, which functionally connect neurons, are at the level of 20-40 nm.

There are several formats to store and share neuron morphologies, including SWC, Eutectic, and Neurolucida. SWC is the most widely used format and it is also embedded in the state-of-the-art neuronal modeling employed in probably the most popular *Neu-*

ICCS Camera Ready Version 2023
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-36021-3_66

roMorpho.Org repository of more than 140,000 neuronal reconstructions. It automatically calculates 21 standard morphometric parameters including the soma surface, number of branches, length, volume, angles, topological asymmetry, fractal dimension, and taper rate [12]. Though this kind of neural modeling is effective for neural tree comparison and cell characterization at the mesoscale, it is not sufficient to fully describe neurons at the nanoscale enabling the determination of synapses and neural circuits. Therefore, a new neuron morphology format at the nanoscale is required.

Our goal here is three-fold, namely, to (1) formulate a complete set of morphologic parameters of the entire neuron at the nanoscale and introduce a new neuronal nanoscale data format; (2) propose four neuronal geometric models based on the introduced neuronal format, and (3) estimate the storage required for the proposed neuronal models.

2 Neuron Morphology and Types

The neuron is the smallest functional processing unit in the nervous system. Neurons are not homogenous and diverse in terms of morphology, connectivity, neuroelectrophysiology, molecular, and genetic properties. Morphologically the neuron comprises a soma (cell body) with two processes (neurites), the dendrites and the axon. The dendrites, which play the role of receptors, receive inputs from other neurons and conduct the impulses to the soma. The axon (or rather the axonal region) acts as a neuron's projector and relays impulses to other cells.

The dendrites form a set of dendritic trunks (stems) each with a dendritic tree. Each tree comprises branches along which dendritic spines with postsynaptic terminals are located. The axonal region contains the hillock which is the soma-axon connector and continues as the axon proper (axonal trunk or stem) terminating as an axonal tree with branches that contain presynaptic terminals.

Neurons are highly variable and no two neurons are the same. Some examples of neuronal types are multipolar (projection and inter) neurons, pyramidal cells, Purkinje cells, or bipolar neurons. It has been evident that there is a plethora of neuronal types and though their total number is still unknown it could be as high as 1000 cell types [13]. Identifying the different brain cell types to determine their roles in health and disease is of great importance [2]. Toward achieving this goal a whole-brain cell atlas is under development [14] that integrates, morphological, physiological, and molecular annotations of neuronal cell types for a comprehensive characterization of cell types, their distributions, and patterns of connectivity.

3 Neuronal Morphologic Parameters and Nanoscale Data Format

To characterize the morphology at the nanoscale, the four groups of parameters are required for the neuron, soma, dendrites, and axon. The proposed neuronal morphologic data format at the nanoscale (nN format) is the following:

ICCS Camera Ready Version 2023
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-36021-3_66

- NEURON
  - Neuron ID
  - SOMA
    - Center coordinates
    - Surface descriptor
  - DENDRITES
    - Number of dendritic trunks
    - For each trunk
      - Trunk ID
      - Proximal coordinates
      - Proximal diameter
      - Dendritic tree root coordinates
      - Dendritic tree root diameter
      - Number of bi(multi)furcations in the dendritic tree
      - Number of terminals in the dendritic tree
      - For each dendritic tree bi(multi)furcation
        - Dendritic tree bi(multi)furcation ID
        - Dendritic tree bi(multi)furcation coordinates
        - Dendritic tree bi(multi)furcation diameter
      - For each dendritic tree terminal
        - Dendritic tree terminal ID
        - Dendritic tree terminal coordinates
        - Dendritic tree terminal diameter
  - AXON
    - Axon ID
    - Hillock proximal (at soma) coordinates
    - Hillock proximal diameter
    - Axonal trunk proximal coordinates
    - Axonal trunk proximal diameter
    - Axonal tree root coordinates
    - Axonal tree root diameter
    - Number of bi(multi)furcations in the axonal tree
    - Number of terminals in the axonal tree
    - For each axonal tree bi(multi)furcation
      - Axonal tree bi(multi)furcation ID
      - Axonal tree bi(multi)furcation coordinates
      - Axonal tree bi(multi)furcation diameter
    - For each axonal tree terminal
      - Axonal tree terminal ID
      - Axonal tree terminal coordinates
      - Axonal tree terminal diameter

ICCS Camera Ready Version 2023
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-36021-3_66

4 Neuronal Morphology Geometric Models

Based on the introduced nano neuron (nN) data format, we consider four neuronal morphology geometric models: straight wireframe, enhanced wireframe, straight polygonal, and enhanced polygonal.

The straight wireframe neuronal model is the simplest. The soma is represented as a center point, the neuronal branches as straight line segments with the start and end points being bifurcations, and the bifurcations and presynaptic and postsynaptic terminals as points.

In the enhanced wireframe neuronal model, in comparison to the straight wireframe one, each branch, besides its start and end points, is determined by additional intermediate points. Hence, a branch forms a polyline segment. Alternatively, the branch points can be connected by the cardinal splines forming a curved branch. With a single intermediate point, the branch is parabolic, and with two points cubic.

The straight polygonal neuronal model, in comparison to the straight wireframe one, requires the knowledge of diameters at the bifurcations and terminals. In this model, the soma, dendrites, and axons are modeled as polygonal surfaces. The soma can have a predefined shape, such as a sphere or pyramid, or a free shape. In the latter case, the soma can be created via iso-surfacing by employing, e.g., the Marching Cubes algorithm [15]. The dendritic and axonal branches are modeled as cylinders or truncated cones.

The enhanced polygonal neuronal model, in comparison to the straight polygonal one, requires the determination of intermediate points considered the centers of cross-sections along with their corresponding diameters. To get a more accurate and better quality of branch surfaces they can be modeled as tubular segments created by subdivision with centerline smoothing and diameter outlier removal [16].

5 Storage Requirements for Volumetric and Geometric Neuronal Morphology Models

Here we consider storage requirements for two classes of neuronal models (1) volumetric with a sampling-dependent size containing the raw (unprocessed) synchrotron tomographic volumetric data, meaning the volumetric data after their reconstruction from projections; and (2) geometric with all complete neurons that have already been extracted from the volumetric data along with their calculated neuronal parameters as specified in the nN format.

5.1 Volumetric Neuronal Models

Let us first consider the case when the average brain of 1,400 cm^3 is acquired with 1 μm^3 spatial resolution with each voxel (sample) of 16-bit intensity resolution. This spatial sampling rate is sufficient to demonstrate the cell bodies. Then, the complete average brain of 1 μm^3 voxels requires $1.4 \times 10^3 \times 10^{12} \times 2 = 2.8 \times 10^{15}$ meaning 2.8 petabytes (PB) of storage.

ICCS Camera Ready Version 2023
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-36021-3_66

However, this 1 μm resolution is not sufficient to fully demonstrate the axons of 1 μm diameter and according to the Nyquist sampling theorem, this sampling resolution shall be no lower than 0.5 μm. Hence, for 0.5 μm spatial sampling resolution, the required storage is increased 8 times to 22.4 PB.

To demonstrate the synapses, which are approximately 20-40 nm wide, the sampling rate shall be at least 10 nm to get the smallest synapses. So, the complete average brain of 10 nm^3 voxels requires $1.4 \times 10^3 \times 10^{18} \times 2 = 2.8 \times 10^{21}$ meaning 2,800 exabytes (EB) of storage.

5.2 Geometric Neuronal Models

For storage estimation, we take the number of neurons of one hundred billion and that of connections of one thousand trillion [17], meaning that each neuron has on average 10,000 connections. We also assume that the neurites form full binary trees.

Let us first consider the straight wireframe neuronal model. To store a neuron's unique identifier out of $10^{11} = 2^{36.5}$ neurons, 5 bytes (B) are needed. To store the soma, 3 center point coordinates are required. Three bytes are needed to store a single value in the range of spatial resolutions from 1 μm ($2^{19.9}$) to 20 nm ($2^{23.6}$). Hence, the soma requires 9 B of storage. Each terminal point has 3 coordinates and requires 9 B of storage. As on average a single neuron has 10,000 terminal points, then all the neuronal terminal points need 90 KB. The number of bifurcations equals the number of terminal points minus one, as in the full binary tree $n = 2l - 1$, where n is the number of nodes and l *is* the number of leaves (i.e., including the terminal points). Each bifurcation has 3 coordinates and needs 9 B. Hence, all bifurcations require approximately 90 KB. Neglecting a small storage requirement for the neuron's identifier in comparison to that for the terminal points and bifurcations, the total storage needed for a single neuron equals approximately 180 KB. Hence, all 10^{11} neurons each of $\sim 180 \times 10^3$ KB require approximately 18 PB of storage.

The enhanced wireframe neuronal model additionally requires intermediate points for neurite branches. Axons may need more intermediate points as they are generally much longer than dendrites. The number of these points is variable and, preferably, shall be curvature-dependent. In the full binary tree the number of branches $b = 2n - 2l$. Since $n = 2l - 1$, then $b = 2l - 2$ meaning that the number of branches approximately doubles that of terminals. Consider the case where each neurite branch has one intermediate point meaning that this branch requires the additional 9 B and is represented by a parabolic segment. Then, the additional storage required for a single neuron is $\sim 2 \times 9$ B $\times 10^4 \approx 180$ KB and 18 PB for the entire brain. Hence, the parabolic wireframe model requires approximately 18 + 18 = 36 PB of storage, the cubic wireframe model 54 PB, and every additional intermediate branch point increases the required storage by 18 PB.

Let us consider the straight polygonal neuronal model. Then, the branches are modeled as cylinders or truncated cones, and let us assume that the soma has a spherical shape. As these regular surfaces with a huge number of polygons can be calculated on demand, there is no need to store the polygonal vertices and the normals. However, the bifurcation and terminal diameters shall be taken into account. Then, to store the soma,

ICCS Camera Ready Version 2023
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-36021-3_66

3 center point coordinates and 1 diameter are required meaning 12 B. Each terminal is characterized by 3 coordinates and 1 diameter, so all the terminals need 120 KB. Each bifurcation is characterized by 3 coordinates and 1 diameter, thus all bifurcations require approximately 120 KB. Neglecting a small storage requirement for the neuron's identifier and soma, the total storage needed for all 10^{11} neurons each of ~240 x 10^3 KB amounts to approximately 24 PB.

In the enhanced polygonal neuronal model a single intermediate cross-section with its center point and diameter requires 12 B, so the additional storage needed for a single neuron is ~2 x 12 B x $10^4 \approx 240$ KB and 24 PB for the entire brain. Hence, the parabolic polygonal model requires approximately 24 + 24 = 48 PB of storage, the cubic wireframe model 72 PB, and every additional intermediate branch cross-section increases the required storage by 24 PB.

6 Discussion

One of the key challenges in neuroscience is to create the nanoscale human connectome. Towards this objective, we here present a complete nanoscale morphometric model of the neuron from which the synapses can subsequently be extracted, the neuronal circuits formed, and the nanoscale connectome produced. To my best knowledge, this is the first work providing for the human brain (1) the complete set of neuronal morphology parameters, (2) neuronal nanoscale data format, and (3) storage requirement estimation for both volumetric and geometric neuronal morphology models at the micro and nanoscales.

Solving neuroscience problems requires high-performance computing and big data [18]. Here we quantitatively address this data bigness in the context of the human nanoscale connectome. The size of the raw volumetric data is spatial resolution-dependent and ranges from 2.8 PB for 1 μm spatial resolution to 2,800 EB for 10 nm resolution.

The size of the straight wireframe neuronal model is approximately 18 PB. The parabolic wireframe model needs approximately 36 PB, the cubic wireframe model 54 PB, and every additional intermediate branch point increases the required storage by 18 PB. The straight polygonal model requires approximately 24 PB. The parabolic polygonal model needs 48 PB, the cubic polygonal model 72 PB, and every additional intermediate branch cross-section increases the required storage by 24 PB. Note that the branches in the parabolic models are planar with a constant curvature sign which limitations are overcome with the cubic models. The sizes of the geometric models, in contrast to those of the volumetric ones, remain constant in the range of the considered here spatial resolutions from 1 μm to 20 nm.

The sizes of the volumetric and geometric models are comparable at the spatial level enabling axon identification whereas at the level of synapses the geometric model hugely outperforms the volumetric model in terms of size. Moreover, the geometric model is easier to automatically process, enhance, extend, mine, visualize, and interact with.

ICCS Camera Ready Version 2023
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-36021-3_66

The future work includes storage estimation for the neuronal circuits and the assessment of high-performance computing resources necessary to process neurons and neural circuits. This in turn may lead to the development of a human brain atlas at macro, meso, micro, and nano scales as proposed in [19].

References

1. Nowinski, W.L.: Advances in neuroanatomy through brain atlasing. Anatomia 2(1):28-42 (2023).
2. BRAIN Initiative BRAIN Working Group. (2014). BRAIN 2025. A Scientific Vision. NIH; https://www.braininitiative.nih.gov/pdf/BRAIN2025_508C.pdf, last accessed 2023/02/17.
3. White, J.G., Southgate, E., Thomson, J.N., Brenner, S.: The structure of the nervous system of the nematode Caenorhabditis elegans. Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences 314(1165), 1–340 (1986).
4. Ryan, K., Lu, Z., Meinertzhagen, I.A.: The CNS connectome of a tadpole larva of Ciona intestinalis (L.) highlights sidedness in the brain of a chordate sibling. Elife 5:e16962, doi: 10.7554/eLife.16962 (2016).
5. Chiang, A.S., Lin, CY, et al.: Three-dimensional reconstruction of brain-wide wiring networks in Drosophila at single-cell resolution. Curr Biol. 21(1), 1-11 (2011).
6. Van Essen, D.C.: Cartography and connectomes. Neuron 80, 775–790 (2013).
7. Landhuis, E.: Neuroscience: Big brain, big data. Nature 541:559–561 (2017).
8. Plaza, S.M., Scheffer, L.K., Chklovskii, D.B.: Toward large-scale connectome reconstructions. Curr Opin Neurobiol 25C, 201-210 (2014).
9. Chin, A-L., Yang, S-M., Chen, H-H., et al.:. A synchrotron X-ray imaging strategy to map large animal brains. Chinese Journal of Physics 65, 24–32 (2020).
10. Hwu, Y., Margaritondo, G., Chiang, A.S. Q&A: Why use synchrotron x-ray tomography for multi-scale connectome mapping? BMC Biol. 15(1), 122. doi: 10.1186/s12915-017--8 (2017).
11. Walsh, C.L., Tafforeau, P., Wagner, W.L., et al.: Imaging intact human organs with local resolution of cellular structures using hierarchical phase-contrast tomography. Nature Methods 18(12), 1532-1541 (2021).
12. Akram, M.A., Ljungquist, B., Ascoli, G.A.: Efficient metadata mining of web-accessible neural morphologies. Progress in Biophysics and Molecular Biology 168, 94-102 (2022).
13. Allen Cell Types Database. Technical white paper: overview. https://help.brain-map.org/display/celltypes/Documentation, last accessed 2023/02/17
14. Ecker, J.R., Geschwind, D.H., Kriegstein, A.R., et al.: The BRAIN Initiative Cell Census Consortium: Lessons learned toward generating a Comprehensive Brain Cell Atlas. Neuron 96(3), 542-557 (2017).
15. Lorensen, W., Cline, H.: Marching cubes: a high resolution 3-D surface construction algorithm. Computer Graphics 21(4), 163-169 (1987).
16. Volkau, I., Zheng, W., Aziz, A., Baimouratov, R., Nowinski, W.L: Geometric modeling of the human normal cerebral arterial system. IEEE Transactions on Medical Imaging 24(4), 529-539 (2005).
17. DeWeerdt, S. How to map the brain. Nature 571, S6-S8 (2019).
18. Chen, S., He, Z., Han, X., et al.: How big data and high-performance computing drive brain science. Genomics, Proteomics & Bioinformatics 17(4), 381-392 (2019).
19. Nowinski, W.L.: Towards an architecture of a multi-purpose, user-extendable reference human brain atlas. Neuroinformatics 20, 405-426 (2022).

ICCS Camera Ready Version 2023
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-36021-3_66