From the synaptome to the connectome: data bigness estimation for the human connectome at the nanoscale

Wieslaw L. Nowinski[1]

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

Abstract. Knowledge of the human nanoscale connectome is crucial for understanding brain function in health and disease. However, the data required to construct a complete nanoscale connectome remain unavailable, and the exact numbers of circuits forming the connectome and neurons within each circuit are still unknown. This study introduces nanoscale morphologic connectomic wireframe and geometric models, each comprising three sub-models (straight and enhanced with parabolic and cubic branches); provides formulas to estimate their data bigness; and assesses required storage. The connectome size/storage estimation builds upon prior work on the synaptome (complete synapse set). To account for the great variability in neuronal and synaptic counts, two estimates for the total number of brain neurons (86 and 100 billion) and three estimates for synapses-per-neuron (1,000;10,000; and 30,000) are considered across six connectomic models, yielding 36 storage estimation cases. The straight wireframe model requires from 8.51PB (for 86 billion neurons, 1,000 synapses-per-neuron) to 297PB (for 100 billion neurons, 30,000 synapses-per-neuron). The straight geometric model needs from 10.58PB (for 86 billion neurons, 1,000 synapses-per-neuron) to 369PB (for 100 billion neurons, 30,000 synapses-per-neuron). Model enhancement significantly increases storage from 22.27PB for the parabolic wireframe model (for 86 billion neurons, 1,000 synapses-per-neuron) to 1,569PB for the cubic geometric model (for 100 billion neurons, 30,000 synapses-per-neuron). The storage required for the complete human nanoscale connectome, as estimated for six models and 36 cases, exceeds the capacity of today's most powerful supercomputers. This work is the first providing the bigness data estimation for representing the entire human nanoscale connectome.

Keywords: Human Brain, Modeling, Storage, Connectome, Synaptome, Neuron, Synapse, Big Data, Nanoscale, Complex systems.

1 Introduction

The human brain is the most complex spatiotemporal system in the known universe. Its spatial scale ranges from nanometers to centimeters, encompassing molecules, synapses, neurons, circuits, nuclei, tracts, lobes, and the entire brain. The temporal scale spans from milliseconds to millions of years, involving processes such as neurotransmission, synaptic connectivity, neural plasticity, brain development, disease progression, aging, and evolutionary changes. Despite significant advancements in neuroscience, there is

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97554-7_21

presently no technology capable of investigating the human brain across all these scales.

To elucidate brain structure and function there has been an enormous global explosion of human brain-related advanced and big projects in the last few years, such as *The BRAIN Initiative* (*Brain Research through Advancing Innovate Neurotechnologies*) to develop technology to advance neuroscience discovery; *The Human Connectome Project* to map structural and functional connections to investigate the relationship between brain circuits and behavior; *The Allen Brain Atlas* to map gene expression; *The Human Brain Project* to create a research infrastructure to decipher the human brain, reconstruct its multiscale organization, and develop brain-inspired technology; *The Big Brain* to acquire ultra-high resolution neuroimages; the *CONNECT* project combining macro- and micro-structure; the *Brainnetome* to understand the brain and its disorders, develop methods for multi-scale brain network analysis, and create the *Brainnetome* atlas; *The Blue Brain Project* to simulate neocortical micro-circuitry; the *Chinese Color Nest Project* to study human connectomics across the life span; the Japanese *Brain/MINDS* (*Brain Mapping by Integrating Neurotechnologies for Disease Studies*) project to better understand the human brain and neuropsychiatric disorders through 'translatable' biomarkers; and *SYNAPSE* (*Synchrotron for Neuroscience – an Asia-Pacific Strategic Enterprise*) to map the entire human brain at sub-cellular level by employing synchrotron tomography [1]. These projects, along with other brain-related efforts, have resulted in the acquisition of big data and the development of diverse brain maps and atlases [2].

The nervous system is made up of neurons that communicate via synapses and the neural connections form a connectome. One of the key challenges in neuroscience is to build the human nanoscale connectome and to develop the corresponding connectomic brain atlas. Knowledge of the connectome is critical for understanding how neural circuits encode information and how the brain functions in health and disease [3]. So far, the full nanoscale connectomes have only been completed for the nematode *Caenorhabditis elegans* (*C.elegans*) [4], larva *Ciona intestinalis* [5], and *Drosophila* [6]. These animals have very small brains; namely, *Ciona intestinalis* 177 neurons and about 6,600 connections, *C.elegans* 302 neurons (a constant number) and about 7,000 connections, and *Drosophila* about 100-130 thousand neurons and millions of connections. The human connectome has only been created at the macroscale by providing anatomical and functional connectivity [7,8]. However, the complete human nanoscale connectome has not yet been developed. What is presently available is the first large, nanoscale human brain tissue sample of 1 mm^3 [9]. This sample was surgically acquired from the temporal lobe, embedded in resin, cut into ~30 nm thick sections, and scanned using electron microscopy, resulting in a volume of 1.4 petabytes of data comprising 57,000 cells and 150 million synapses.

The most critical obstacles in mapping the complete human nanoscale connectome are the prohibitively long overall time necessary to acquire data and the immense computational resources required for storing and processing such huge data [10,11,12]. For instance, imaging the entire human brain at the nanoscale resolution is estimated to require 17 million years [11] when applying the same brain imaging protocol as was employed for *Drosophila* [6]. Fortunately, due to the progress in imaging, the acquisition time continually decreases. In particular, synchrotron X-ray tomography is a promising imaging modality offering to decrease the whole human brain acquisition time at

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97554-7_21

the sub-cellular level to a few years or even less [13]. Its advantages include nanoscale resolution, high-depth penetration, whole-brain imaging without sectioning, and high-speed 3D imaging.

In anticipation of the future availability of a whole human brain nanoscale volume suitable for processing, modeling, and atlasing, my current efforts focus on the estimation of computational resources needed for undertaking this monumental endeavor. At the ICCS 2023 conference, I proposed a dedicated nano neuronal (*nN*) data file format to describe neuron morphology at the nanoscale [14]. This format comprises the complete dendritic and axonal trees without any simplification or reductionist encoding and includes the dendritic and axonal terminals determining the location and size of synapses. The *nN* format was subsequently extended to embed gross neuroanatomy to facilitate atlas-enabled navigation and exploration of the brain model at the nanoscale [15]. At the ICCS 2023, I also introduced four geometric neuronal morphology models at the nanoscale: straight wireframe, enhanced wireframe, straight polygonal, and enhanced polygonal. Subsequently, I estimated storage requirements for these geometric models and additionally a volumetric neuronal model [16]. Then, the human nanoscale synaptome morphology was modeled and its storage requirements were estimated [17].

This study stems from these previous works and tackles the problem of big data estimation in the human connectome at the nanoscale. The problem is challenging not only because such data are not yet available but also because both the exact number of circuits forming the connectome and the number of neurons within each circuit remain unknown. I tackle this problem here by deriving the connectome storage estimation from the previously computed estimates for the synaptome [17].

This work aims to 1) introduce two groups of nanoscale morphology connectomic models: wireframe (skeletal) by combining the synaptic point model [17] with the neuronal wireframe model [16] and geometric by merging the synaptic geometric model [17] with the neuronal polygonal model [16], each group comprising 3 sub-groups (straight and enhanced with parabolic and cubic branches); 2) provide formulas for data bigness estimation of the human nanoscale connectome for various connectomic models; and 3) estimate storage requirements for these connectomic models.

To account for a wide range of variations in the numbers of neurons and synapses reported in the literature, two cases of the total number of neurons in the human brain (86 and 100 billion) and three cases of the average number of synapses per neuron (1,000;10,000; and 30,000) are considered across six connectomic models resulting in total 36 cases of storage estimation. To the best of my knowledge, this is the first comprehensive big data study providing a quantitative estimation of the storage requirements for the human nanoscale connectome.

2 Method and material

2.1 Method

A neuron consists of a soma (cell body) and its projections, known as neurites (neuronal processes), which include dendrites and a single axon [18]. The dendrites receive impulses from other neurons and transmit them to the soma. The axon acts as a neuron's projector and relays impulses to other cells. The dendrites form a set of dendritic trunks,

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97554-7_21

each with a dendritic tree. Each tree comprises branches along which the dendritic spines with postsynaptic terminals (postsynaptic densities) are located. The proximal segment of the axonal neurite contains the hillock, which is the soma-axon connector, and distally continues as the axon proper terminating as an arborized axonal tree with multiple branches comprising presynaptic terminals (presynaptic active zones or densities). The neurons connected through the synapses form neuronal circuits and the entirety of these circuits constitutes the connectome.

A direct approach of calculating the size of the connectomic big data and the corresponding storage required for keeping them is to execute the following algorithm

For every neuronal circuit
  For every neuron in the neuronal circuit
    Calculate storage for the neuron
    Accumulate neurons' storage

The problem with this algorithm is that the numbers of neuronal circuits, neurons within each circuit, dendritic trees within each neuron, and branches within the dendritic and axonal trees remain unavailable even for the average human brain.

The approach I propose here is to estimate the connectomic big data from the synaptome. The synaptome is the set of all the synapses. The connectome is formed by the synaptome and all its presynaptic and postsynaptic neurons without their presynaptic and postsynaptic terminals. Each connecting neuron comprises the soma, the dendritic trunks with the trees without their postsynaptic terminals, and the hillock and axon with the axonal tree without its presynaptic terminals. Then, the connectome is defined as

$$\begin{aligned} Connectome = Synaptome \\ + \sum_{\substack{Neurons\ forming \\ connectome}} (soma + dendritic\ trunks\ with\ trees \\ + hillock, axon\ with\ axonal\ tree) \end{aligned} \tag{1}$$

Formally, the above summation is over all the presynaptic and postsynaptic neurons. However, since each neuron is likely to participate in multiple circuits, summing over all neurons is sufficient, thereby avoiding duplication in neuron counting. Hence,

$$\begin{aligned} Connectome = Synaptome \\ + \sum_{\substack{All \\ neurons}} (soma + dendritic\ trunks\ with\ trees + hillock, axon\ with\ axonal\ tree) \end{aligned} \tag{2}$$

The human nanoscale synaptome was modeled and its data bigness estimated earlier [17]. Namely, a synapse is formed by two neurons meaning the following pair of pairs [(presynaptic neuron),(presynaptic axonal terminal);(postsynaptic neuron),(postsynaptic dendritic terminal)]. Let i and j be a presynaptic neuron and a postsynaptic neuron with identifiers N_i and N_j, and m and n the axonal terminal A_{im} and the dendritic terminal D_{jn} of neurons i and j, respectively. Moreover, let (x,y,z) determine terminal coordinates

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97554-7_21

and r terminal radius. Then, the synapse between the neurons i and j is defined as a pair of quadruples $[N_i,A_{im},(x_{im},y_{im},z_{im}),r_{im};N_j,D_{jn},(x_{jn},y_{jn},z_{jn}),r_{jn}]$. A synapse has three components, topology (the neuron and terminal identifiers), location (the coordinates of the terminal centers), and geometry (the terminal radii by approximating the fundi of the terminals with the circular shape). Consequently, I earlier introduced three synaptic morphology models with diverse sizes, content, and potential applications; namely, the topologic model with topology, the point model with topology and location, and the geometric model with topology, location, and geometry [17]. The complete synaptome is defined as a set of all the synapses $\{[N_i,A_{im},(x_{im},y_{im},z_{im}),r_{im};N_j,D_{jn},(x_{jn},y_{jn},z_{jn}),r_{jn}]\}$ for all the presynaptic and the corresponding postsynaptic neurons.

I also proposed four neuronal geometric models: straight wireframe, enhanced wireframe, straight polygonal, and enhanced polygonal [16]. The straight wireframe neuronal model is the simplest with the soma represented as a center point, the neuronal branches as straight-line segments with the start and end points being bifurcations, and the presynaptic and postsynaptic terminals as points. The enhanced wireframe neuronal model is obtained from the straight wireframe model by using additional intermediate points placed on neurite branches to enhance their shape. Then, a branch forms a polyline segment; alternatively, by connecting the branch points by cardinal splines a curved branch is obtained. A single intermediate point results in a parabolic branch and two points produce a cubic branch. In the straight polygonal neuronal, 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. The dendritic and axonal branches are modeled as cylinders or truncated cones, so this tubular model requires the knowledge of radii at the bifurcations and terminals. Finally, the enhanced polygonal neuronal model requires the determination of intermediate points along each branch treated as the centers of cross-sections, each associated with a corresponding radius.

Let us estimate the data necessary to handle a single neuron without its terminals in the straight wireframe neuronal model. In general, the data size is spatial resolution-dependent, and the data sampling resolution shall be sufficient to distinguish the synapses. There are two types of synapses: chemical with the synaptic gap (cleft) of 20-30/20-40 nm [19,20] and electrical with the 2-4 nm gap [19]. To distinguish the synapses, the acquired data shall be sampled sufficiently dense. Propel sampling is critical since axonal terminals (boutons) are adjacent to several possible synaptic targets [21]. According to the Nyquist sampling theorem, the spatial sampling resolution shall be no lower than 10 nm to handle the chemical synapses and 1 nm for the electrical synapses. Then, for the spatial resolutions either of 10 nm ($10^8 = 2^{26.6}$) or 1 nm ($10^9 = 2^{29.9}$), 4B are required. To store the soma, its 3 center point coordinates are needed, meaning 12 B, which can be neglected in comparison to big data of the dendritic and axonal threes.

Let us assume that the dendritic and axonal neurites form perfect binary trees. In the perfect binary tree $n = 2l - 1$, where n is the number of nodes (including the leaves) and l *is* the number of leaves. Hence, the number of bifurcations for a high number of terminals approximately equals that of terminals. Moreover, every node in the perfect binary tree, excluding the terminals, generates two branches b, so $b = 2(n - l)$. Therefore, taking into account that $n = 2l - 1$, the number of branches *is* $b = 2l - 2$, meaning that for a high number of terminals, the number of branches approximately doubles that of terminals. The *nN* format does not store the branches, but only their endpoints (nodes),

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97554-7_21

each with 3 coordinates and an index. Three coordinates require 12 B and an index of 2 B is sufficient to distinguish up to 65,000 objects (which number is smaller than the number of synapses per neuron reported in the literature, see below). Moreover, the additional 2B are required to provide the index of a neighboring node in the neuronal tree to which the considered node is connected. Therefore, a single node needs 16 B of data without considering its geometry. By including geometry, i.e., a node radius, 20 B of data are required per node. As the number of nodes approximately equals that of terminals, so for the nanoscale resolutions of 1 nm and 10 nm, the wireframe connectome, i.e., that with node and terminal locations but without its geometry, approximately requires the following amount of data

$$Connectome\_wireframe = Synaptome + 16 * Total\_number\_of\_synapses \tag{3}$$

and when including geometry

$$Connectome\_geometric = Synaptome + 20 * Total\_number\_of\_synapses \tag{4}$$

Note that "Synaptome" in Eq. 3 corresponds to the point synaptic model and that in Eq. 4 to the geometric synaptic model. The total number of synapses (*TNS*) can be approximated as

$$Total\_number\_of\_synapses = 0.5 * Number\_of\_neurons * Average\_number\_of\_synapses\_per\_neuron \tag{5}$$

(note that as a synapse is formed by two neurons, half of the total number of neurons is taken in Eq. 5 to avoid synapse duplication).

The neuronal model enhancement consists in the insertion of additional intermediate points (or cross-sections for the geometrical tubular models) at the neurite branches [16]. One additional point along with the indices of the endpoints of the branch it belongs to requires the storage of (12 + 4) B x 2 x number of terminals (synapses per neuron). Hence,

$$Connectome\_enhanced\_wireframe\_parabolic = Synaptome + (16 + 32) * Total\_number\_of\_synapses \tag{6}$$

and two additional points double the amount of these data

$$Connectome\_enhanced\_wireframe\_cubic = Synaptome + (16 + 2 * 32) * Total\_number\_of\_synapses \tag{7}$$

The enhanced geometric model requires one additional point with two indices and the cross-section radius resulting in the amount of (16 + 4) B x 2 x number of terminals data

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97554-7_21

$$Connectome\_enhanced\_geometric\_parabolic = Synaptome + (20 + 40) * Total\_number\_of\_synapses \quad (8)$$

and two additional points double the amount of these data

$$Connectome\_enhanced\_geometric\_cubic = Synaptome + (20 + 2 * 40) * Total\_number\_of\_synapses \quad (9)$$

2.2 Material

The numbers of neurons, all synapses, and synapses per neuron required in Eqs. (3) – (9) are taken here from the literature studies. Because of the immense complexity and variability of the human brain, the exact numbers of both its neurons and synapses are challenging to precisely determine. For instance, the number of neurons in the entire human brain varies from 30 billion [22] to 125 billion [23]. Numerous textbooks, original articles, and reviews give the total number of neurons in the human brain one hundred billion [24-29]. This affirmation is also present in the textbook by Kandel et al. [25], a commonly used textbook in neuroscience authored by a Nobel Prize laureate in physiology or medicine in the year 2000. This textbook knowledge has been challenged by more robust methods of neuron counting providing an estimate of 86.1 billion neurons [30].

Therefore, for the storage estimation, we take 2 values, 86 billion and 100 billion neurons in the whole human brain.

The number of synapses per neuron is also highly variable. Kandel et al. estimate that there are around 100 trillion synapses in the average adult human brain [25] (meaning 1,000 synapses per neuron). According to a reference book [31], the average neuron has 1,000 synapses with other neurons. Ten thousand connections per neuron are reported in [28]. Some sources estimate that a single neuron can have between 1,000–15,000 synaptic connections [32]. DeFelipe demonstrated that the number of synapses per neuron is cortical layer-dependent with the average number of synapses per neuron being 29,642 (100,042 in layer I, 17,046 in layer II, 37,066 in layer IIIa, 56,521 in layer IIIb, 15,989 in layer IV, 29,965 in layer V, and 28,224 in layer VI) [33].

Hence, for the storage estimation, we take 3 average values, 1,000, 10,000, and 30,000 synapses per neuron.

3 Results

The storage required for the human connectome at the nanoscale is estimated for 5 neuroanatomical parameters (86 and 100 billion neurons and 1,000, 10,000 and 30,000 average synapses per neuron) and 2 groups of connectomic models (wireframe and geometric) with 3 models each (straight and enhanced parabolic and cubic) resulting in 36 cases of storage estimation.

Let us consider the straight wireframe and geometric connectomic models for 100 billion neurons. For 1,000 synapses per neuron, the point synaptic model requires 1.9

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97554-7_21

PB and the geometric synaptic model 2.3 PB [17]. The total number of synapses *TNS* = 100 billion x 0.5 x 1,000 = 500 billion (0.5 PB), see (Eq. 5). By applying Eq. 3, the straight wireframe connectomic model requires 1.9 PB + 16 x 0.5 PB = 9.9 PB and from Eq. 4 the straight geometric connectomic model needs 2.3 PB + 20 x 0.5 PB = 12.3 PB. For 10,000 synapses per neuron, the point synaptic model requires 19 PB and the geometric synaptic model needs 23 PB, whereas for 30,000 synapses per neuron the point synaptic model requires 57 PB and the geometric synaptic model needs 69 PB [17]. The *TNS* equals 5 PB and 15 PB for 10,000 and 30,000 synapses per neuron, respectively. Then, for 10,000 synapses per neuron, the straight wireframe connectomic model requires 19 PB + 16 x 5 PB = 99 PB and the straight geometric connectomic model needs 23 PB + 20 x 5 PB = 123 PB. For 30,000 synapses per neuron, the straight wireframe connectomic model requires 57 PB + 16 x 15 PB = 297 PB and the straight geometric connectomic model needs 69 PB + 20 x 15 PB = 369 PB.

Let us consider the enhanced connectomic models for 100 billion neurons. For 1,000 synapses per neuron, the enhanced wireframe connectome requires 1.9 PB + 48 x 0.5 PB = 25.9 PB for the parabolic model (Eq. 6) and 1.9 PB + 80 x 0.5 PB = 41.9 PB for the cubic model (Eq. 7). Whereas the enhanced geometric connectome needs 2.3 PB + 60 x 0.5 PB = 32.3 PB for the parabolic model (Eq. 8) and 2.3 PB + 100 x 0.5 PB = 52.3 PB for the cubic model (Eq. 9). For 10,000 synapses per neuron, the enhanced wireframe connectome requires 19 PB + 48 x 5 PB = 259 PB for the parabolic model (Eq. 6) and 19 PB + 80 x 5 PB = 419 PB for the cubic model (Eq. 7). Whereas the enhanced geometric connectome needs 23 PB + 60 x 5 PB = 323 PB for the parabolic model (Eq. 8) and 23 PB + 100 x 5 PB = 523 PB for the cubic model (Eq. 9). For 30,000 synapses per neuron, the enhanced wireframe connectome requires 57 PB + 48 x 15 PB = 777 PB for the parabolic model (Eq. 6) and 57 PB + 80 x 15 PB = 1,257 PB for the cubic model (Eq. 7). Whereas the enhanced geometric connectome needs 69 PB + 60 x 15 PB = 969 PB for the parabolic model (Eq. 8) and 69 PB + 100 x 15 PB = 1,569 PB for the cubic model (Eq. 9).

Let us consider the cases corresponding to 86 billion neurons. Then, the point synaptic model requires 1.63 PB, 16.34 PB, and 49.02 PB for 1,000, 10,000, and 30,000 synapses per neuron, respectively; whereas the geometric synaptic model needs 1.98 PB, 19.78 PB, and 59.34 PB for 1,000, 10,000, and 30,000 synapses per neuron, respectively [17]. The *TNS* equals 86 billion x 0.5 x 1,000 = 430 billion corresponding to 0.43 PB for 1,000 synapses per neuron, 4.3 PB for 10,000 synapses per neuron, and 12.9 PB for 30,000 synapses per neuron. Then, for 1,000 synapses per neuron, the straight wireframe connectomic model requires 1.63 PB + 16 x 0.43 PB = 8.51 PB whereas the straight geometric connectomic model needs 1.98 PB + 20 x 0.43 PB = 10.58 PB. For 10,000 synapses per neuron, the straight wireframe connectomic model requires 16.34 PB + 16 x 4.3 PB = 85.14 PB whereas the straight geometric connectomic model needs 19.78 PB + 20 x 4.3 PB = 105.78 PB. For 30,000 synapses per neuron, the straight wireframe connectomic model requires 49.02 PB + 16 x 12.9 PB = 255.42 PB whereas the straight geometric connectomic model needs 59.34 PB + 20 x 12.9 PB = 317.34 PB.

Consider the enhanced connectomic models for 86 billion neurons. For 1,000 synapses per neuron, the enhanced wireframe connectome requires 1.63 PB + 48 x 0.43 PB = 22.27 PB for the parabolic model (Eq. 6) and 1.63 PB + 80 x 0.43 PB = 36.03 PB for the cubic model (Eq. 7). Whereas the enhanced geometric connectome needs 1.98

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97554-7_21

PB + 60 x 0.43 PB = 27.78 PB for the parabolic model (Eq. 8) and 1.98 PB + 100 x 0.43 PB = 44.98 PB for the cubic model (Eq. 9). For 10,000 synapses per neuron, the enhanced wireframe connectome requires 16.34 PB + 48 x 4.3 PB = 222.74 PB for the parabolic model (Eq. 6) and 16.34 PB + 80 x 4.3 PB = 360.34 PB for the cubic model (Eq. 7). Whereas the enhanced geometric connectome needs 19.78 PB + 60 x 4.3 PB = 277.78 PB for the parabolic model (Eq. 8) and 19.78 PB + 100 x 4.3 PB = 449.78 PB for the cubic model (Eq. 9). For 30,000 synapses per neuron, the enhanced wireframe connectome requires 49.02 PB + 48 x 12.9 PB = 668.22 PB for the parabolic model (Eq. 6) and 49.02 PB + 80 x 12.9 PB = 1,081.02 PB for the cubic model (Eq. 7). Whereas the enhanced geometric connectome needs 59.34 PB + 60 x 12.9 PB = 833.34 PB for the parabolic model (Eq. 8) and 59.34 PB + 100 x 12.9 PB = 1,349.34 PB for the cubic model (Eq. 9).

These results are summarized in Table 1 and are also visually presented in Figure 1.

Table 1. Storage in PB required for the complete human nanoscale connectome for the given number of neurons (86 and 100 billion) and synapses per neuron (1,000; 10,000; and 30,000) estimated for the wireframe and geometric groups of connectomic models (each comprising straight, parabolic, and cubic models).

Type of connectomic model		Number of neurons and synapses per neuron					
		1,000		10,000		30,000	
		86 billion	100 billion	86 billion	100 billion	86 billion	100 billion
Straight wireframe		8.51	9.90	85.14	99.00	255.42	297.00
Enhanced wireframe	Parabolic	22.27	25.90	222.74	259.00	668.22	777.00
	Cubic	36.03	41.90	360.34	419.00	1,081.02	1,257.00
Straight geometric		10.58	12.30	105.78	123.00	317.34	369.00
Enhanced geometric	Parabolic	27.78	32.30	277.78	323.00	833.34	969.00
	Cubic	44.98	52.30	449.78	523.00	1,349.34	1,569.00

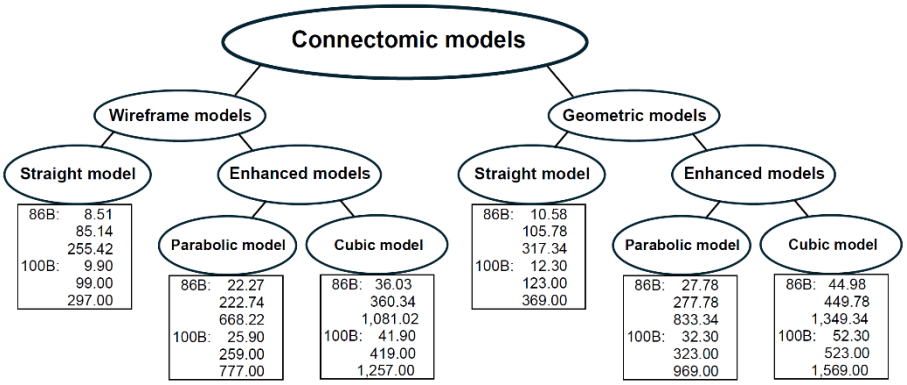

Fig. 1. The results summarized diagrammatically, where the ellipses denote the connectomic models and the rectangles comprise the model data size in PB (the first 3 rows correspond to 1,000, 10,000 and 30,000 synapses per neuron for 86 billion neurons (86B) and the next 3 rows for 100 billion neurons (100B)).

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97554-7_21

4 Discussion

One of the primary challenges in connectomics today is the continuous advancement of cutting-edge technology enabling the acquisition of increasingly high-quality connectomic data with the ever-growing spatial resolution, spanning from macroscopic (using, e.g., magnetic resonance imaging) to mesoscopic (applying, e.g., optical microscopy) to microscopic (employing, e.g., electron microscopy) and to nanoscopic (through, e.g., volume electron microscopy); ranging from very small brains like these of *C.elegans* and *Drosophila* to those of rodents, non-human primates, and human; and from small tissue samples to the whole brain. Moreover, advanced imaging techniques are integrated with molecular, genetic, and physiological approaches to enhance synaptic contact identification in order to understand synaptic organization and connectivity [33]. These developments grow the field of nanoneuroanatomy, leading to the creation of diverse nanomaps and preliminary nanoscale brain atlases. Although a complete human brain dataset at the nanoscale resolution remains yet to be acquired (and only a 1 mm^3 sample is available today), such data are likely to be available in the next few (perhaps even three) years.

My contribution to these nanoneuroanatomy efforts is from a computational standpoint. I have earlier proposed [14], extended [15], and applied [16,17] the nano neuronal (*nN*) data file format to describe neuron morphology at the nanoscale; introduced a volumetric and four geometric neuronal nanoscale morphology models (straight wireframe, enhanced wireframe, straight polygonal, and enhanced polygonal) and estimated storage requirements for them [14,16]; designed a nanoscale human brain atlas for the brain model exploration at the nanoscale [15]; estimated storage requirements for the human nanoscale synaptome [17]; and assessed high-performance computing resources for morphology modeling of the entire human brain at the nanoscale [34].

This work advances my previous nanoneuroanatomy-related efforts. In this paper, despite the lack of precise data on the number of neuronal circuits forming the human connectome, the number of neurons within each circuit, the number of dendritic trees per neuron, and the number of dendritic and axonal tree terminals per neuron, I propose a method to estimate the storage required for the connectome leveraging the previous storage assessment of the synaptome. This is feasible by employing formulas (3) - (9) because the synaptic model applied contains the indices of its corresponding presynaptic and postsynaptic neurons. Here I combine the synaptic point model with the neuronal wireframe models, resulting in the connectomic wireframe (skeletal) models, and the synaptic geometric model with the neuronal polygonal models yielding the connectomic geometric models.

Because of great biological variability across individual human brains and different brain regions as well as the diverse cell and synapse counting methods used, the reported numbers of both neurons and synapses per neuron in the human brain and its parts vary widely across literature studies. To cope with this variability, two cases of the total number of neurons in the human brain (86 and 100 billion) and three cases of the average number of synapses per neuron (1,000;10,000; and 30,000) are considered. In addition, I introduce six connectomic models forming two groups: wireframe and geometric models. These models are potentially useful for various morphometric and network-related analyses. The straight wireframe model is the simplest and the most compact. The enhanced models offer more realistic shapes of neurite branches, whereas

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97554-7_21

the geometric models provide surface representations enhancing visualization at the expense of a substantial increase in the overall size of connectomic data.

The simplest connectomic model, the straight wireframe model, requires from 8.51 PB for 86 billion neurons and 1,000 synapses per neuron to 297 PB for 100 billion neurons and 30,000 synapses per neuron. The straight geometric connectomic model needs from 10.58 PB for 86 billion neurons and 1,000 synapses per neuron to 369 PB for 100 billion neurons and 30,000 synapses per neuron. Model enhancement substantially increases the required storage from 22.27 PB for the parabolic wireframe connectomic model with 86 billion neurons and 1,000 synapses per neuron to 1,569 PB for the cubic geometric connectomic model with 100 billion neurons and 30,000 synapses per neuron. Note that the geometric connectomic models do not account for the data needed to store the vertices and normals of the polygons, typically triangles, forming the surface as these polygons can be calculated on the fly based on the model parameters.

The estimated storage capacity required to represent the human connectome at the nanoscale exceeds the capabilities of even the most advanced supercomputers available today. Namely, the world's first exascale and today's most powerful supercomputer *Frontier* [41], with 9,408 computing nodes each with 4 TB of flash memory offering an overall 37 PB memory, is able to handle the straight and parabolic wireframe and geometric connectomic models for 86 and 100 billion neurons and the cubic wireframe connectomic model for 86 billion neurons provided that the models are limited to 1,000 synapses per neuron. Another exascale supercomputer *Aurora* with an aggregate system memory of 10.9 PB [42] can only handle the straight wireframe connectomic model for 1,000 synapses per neuron.

Although the storage capacity of today's most powerful supercomputers remains insufficient to handle the human connectome at the nanoscale, this limitation can be overcome by leveraging either a network of powerful supercomputers enabling to scale computing resources, like *BRAINS* [43], or cloud storage offered by open-source services, such as the *Brain Observatory Storage Service and Database*, for storing and accessing petascale image datasets [44].

It should be emphasized that this work presents an estimation of data bigness for the human nanoscale connectome, since the determination of the precise amount of data is not possible today as a suitable technology enabling the acquisition of the full human brain data at the nanoscale in a reasonable time is not yet available. Moreover, the human brain is highly variable in cytoarchitecture, neurochemistry, connectivity, and functional characteristics across different brain regions. Several studies have highlighted the interindividual variability both in the structural and functional organization of the human brain [35-37]. Functional and structural studies examining macroscopic connectivity in the human cerebral cortex have implied that the high-order associative cortex exhibits greater connectivity compared to the primary cortex, including the somatosensory, motor, and visual cortices [38-39]. Although complete data for the entire human brain are not yet available, a mouse study, such as [40], provides some indications about regional variability. Namely, this study has found (*i*) the highest densities of synapses in the isocortex, olfactory areas, and hippocampal formation; (*ii*) low densities in the pallidum, hypothalamus, brainstem, and cerebellum; (*iii*) a wide range of synapse densities in the striatum and thalamus; (*iv*) homogeneously low densities in the cerebellar cortex; and (*v*) the lowest synaptic densities in the pallidum, hypothalamus, and brainstem.

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97554-7_21

In this work, I employ geometric and wireframe connectome modeling. Note that the difference in the data bigness between the volumetric neuronal model (or raw volumetric data) and the geometric model is dramatic at the nanoscale. Namely, for the spatial 10 nm (chemical synapse) resolution the geometric neuronal models, depending on their type, range from 24 PB to 96 PB, whereas the volumetric model requires about 5.6 ZB (zetta (10^{21}) bytes); at 1 nm (electric synapse) resolution, the geometric neuronal models need similar storage, while the volumetric model requires about 5,600 ZB [16].

The synaptic geometric model employed in this study is relatively simple. It assumes a circular shape of each of the presynaptic and postsynaptic terminals with three coordinates of its center and a radius (which model requires 2 x 16 B per synapse without counting the indices). However, from a geometric perspective, an actual synapse exhibits a distinct size and shape, thereby necessitating more detailed data, especially that to yield accurate insights into the connectivity, it is critical to account for all morphological characteristics of synapses [45]. For instance, it has been postulated that synaptic size correlates with release probability, synaptic strength, efficacy, and plasticity [46]. In general, synapses can be excitatory and inhibitory. The excitatory synapses are asymmetric and the inhibitory synapses are symmetric. The identification of the asymmetric and symmetric synapses is based on the thickness of the postsynaptic density (PSD). The synapses are also classified into four main categories taking into account the shape of the PSD; namely, macular (disk-shaped PSD), perforated (with one or more holes in the PSD), horseshoe-shaped (with an indentation), and fragmented (disk-shaped PSDs with no connection between them, consisting of two or more physically discontinuous PSDs) [47]. Another geometric feature used when studying the synaptic properties is the position of centroids of synaptic junctions [45]. Moreover, postsynaptic targets can be classified as spines (corresponding to axospinous synapses) or dendritic shafts (axodendritic synapses) [45]. Thus, while the type of synapse along with its postsynaptic target only requires the additional 4 bits to capture the abovementioned features and 12 B for the position of the synaptic centroid, however, to store the shapes of synapses and PSDs can increase the required synaptic storage a few times. Additionally, the PSDs vary not only in size but also in composition across brain regions.

5 Conclusions

Despite the lack of precise connectomic data, I proposed a method to estimate storage required for the human nanoscale connectome based on previously formulated synaptic and neuronal models. The wireframe and geometric connectomic models were formulated, each comprising 3 sub-models: straight and enhanced parabolic and cubic. To account for a substantial neuronal and synaptic variability, the storage demands for each connectomic model was estimated for 86 and 100 billion neurons, and 1,000, 10,000 and 30,000 average synapses per neuron resulting in 36 cases. This storage ranges from 8.51 PB (for the straight wireframe model, 86 billion neurons, and 1,000 synapses per neuron) to 1,569 PB (for the cubic geometric model, 100 billion neurons, and 30,000 synapses per neuron) surpassing the capabilities of today's most powerful supercomputers. Moreover, model enhancement significantly increases storage demands. This study is the first providing the bigness data estimation for the complete human nanoscale connectome.

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97554-7_21

Future challenges include the data bigness estimation for connectomic electrical and molecular models, which are anticipated to be substantially larger than the morphologic models, and evaluating the high-performance computing resources required for the computation and analysis of the human nanoscale connectome.

Acknowledgments. This publication is supported by the European Union's Horizon 2020 research and innovation programme under grant agreement Sano No. 857533. This publication is supported by Sano project carried out within the International Research Agendas programme of the Foundation for Polish Science, co-financed by the European Union under the European Regional Development Fund.

The publication was created within the project of the Ministry of Science and Higher Education "Support for the activity of Centers of Excellence established in Poland under Horizon 2020" on the basis of the contract number MEiN/2023/DIR/3796.

I would like to thank Prof. Javier DeFelipe of the Cajal Institute for providing valuable pointers regarding synapses.

Disclosure of Interests. The author has no competing interests to declare that are relevant to the content of this article.

References

1. Nowinski, W.L.: Advances in neuroanatomy through brain atlasing. Anatomia **2**(1), 28-42 https://doi.org/10.3390/anatomia2010004 (2023)
2. Nowinski, W.L.: Evolution of human brain atlases in terms of content, applications, functionality, and availability. Neuroinformatics **19**(1), 1-22 (2021)
3. BRAIN Initiative BRAIN Working Group. BRAIN 2025. A Scientific Vision. NIH (2014); https://www.braininitiative.nih.gov/pdf/BRAIN2025_508C.pdf, last accessed 2024/12/11 (2014)
4. 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)
5. 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 (2016)
6. Chiang, A.S., Lin, CY, Chuang, C.C., Chang, H.M., Hsieh, C.H., Yeh, C.W., Shih, C.T., Wu, J.J. et al.: Three-dimensional reconstruction of brain-wide wiring networks in Drosophila at single-cell resolution. Curr Biol **21**(1); 1-11, doi: 10.1016/j.cub.2010.11.056 (2011)
7. Van Essen, D.C.: Cartography and connectomes. Neuron **80**; 775–790 (2013)
8. Coleman, C., Van Horn, J.D.: Towards comprehensive connectivity modeling. Neuroinformatics **22**, 225–227; https://doi.org/10.1007/s12021-024-09676-4 (2024)
9. Shapson-Coe, A., Januszewski, M., Berger, D.R., Pope, A., Wu, Y., Blakely, T., Schalek, R.L., Li, P.H., Wang, S., Maitin-Shepard, J., Karlupia, N., et al.: A petavoxel fragment of human cerebral cortex reconstructed at nanoscale resolution. Science **384**(6696), eadk4858. doi: 10.1126/science.adk4858 (2024)
10. Chen, S., He, Z., Han, X., He, X., Li, R., Zhu, H., Zhao, D., Dai, C., Zhang, Y., Lu, Z., Chi, X., Niu, B:. How big data and high-performance computing drive brain science. Genomics, Proteomics & Bioinformatics **17**(4), 381-392 (2019)

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97554-7_21

11. Lichtman, J., Pfister, H. & Shavit, N.: The big data challenges of connectomics. Nat Neurosci **17**, 1448–1454. https://doi.org/10.1038/nn.3837 (2014)
12. Landhuis, E.: Neuroscience: Big brain, big data. Nature **541**, 559–561 (2017)
13. Chin, A-L., Yang, S-M., Chen, H-H., Li, M-T. et al.: A synchrotron X-ray imaging strategy to map large animal brains. Chinese Journal of Physics **65**:24–32 (2020)
14. Nowinski, W.L.: Toward the human nanoscale connectome: neuronal morphology format, modeling, and storage requirement estimation. Proc. International Conference on Computational Sciences ICCS 2023, Prague, Czechia, 3-5 July 2023. Lecture Notes in Computer Science vol. 14074, Part II, pp. 691-698. Springer, Heidelberg. DOI: 10.1007/978-3-031-36021-3_66 (2023)
15. Nowinski, W.L.: Toward morphologic atlasing of the human whole brain at the nanoscale. Big Data and Cognitive Computing **7**(4), 179; https://doi.org/10.3390/bdcc7040179 (2023)
16. Nowinski, W.L.: Storage estimation in morphology modeling of the human whole brain at the nanoscale. Journal of Computational Science **81**, 102346. https://doi.org/10.1016/j.jocs.2024.102346 (2024)
17. Nowinski, W.L.: On human nanoscale synaptome: morphology modeling and storage estimation. PLOS ONE, September 25, 2024; https://doi.org/10.1371/journal.pone.0310156 (2024)
18. Mumenthaler, M., Mattle H.: Fundamentals of Neurology: An Illustrated Guide. Thieme, Stuttgart – New York (2005)
19. Citow, J.S., Macdonald, R.L.: Neuroanatomy and Neurophysiology: a Review. Thieme, New York (2001)
20. Patestas M.A, Gartner L.P.: A Textbook of Neuroanatomy. Blackwell Publishing, Malden MA (2006)
21. Turegano-Lopez, M., de las Pozas, F., Santuy, A. et al.: Tracing nerve fibers with volume electron microscopy to quantitatively analyze brain connectivity. Commun Biol **7**, 796. https://doi.org/10.1038/s42003-024-06491-0 (2024)
22. Szentágothai J.: The modular architectonic principle of neural centers. Rev Physiol Biochem Pharmacol. **98**, 11–61 (1983)
23. Lent R., Azevedo F.A., Andrade-Moraes C.H., Pinto A.V.: How many neurons do you have? Some dogmas of quantitative neuroscience under revision. Eur J Neurosci. **35**(1), 1-9. doi: 10.1111/j.1460-9568.2011.07923.x (2012)
24. Fields, RD.: The Other Brain. Simon & Schuster, New York (2009)
25. Kandel, E. R., Schwartz, J. H., and Jessell, T. M.: Principles of Neural Science. 4th ed. McGraw-Hill, New York (2000)
26. Kandel, E.R.; Schwartz, J.H.; Jessell, T.M.; Siegelbaum, S.A.; Hudspeth, AJ. Principles of Neural Science. 5th edn. McGraw-Hill, New York (2013)
27. Nolte, J.: The Human Brain. 3rd edn. The C.V. Mosby Company, St. Louis (1993)
28. Purves, D.: Brains: How they work and what that tells us about who we are. FT Press, Upper Saddle River (2010)
29. DeWeerdt, S.: How to map the brain. Nature **571**, S6-S8 (2019)
30. Azevedo, F.A., Andrade-Moraes, C.H., Curado, M.R., Oliveira-Pinto, A.V., Guimarães, D.M., Szczupak, D., Gomes, B.V., Alho, A.T., Polichiso, L., Tampellini, E., Lima, L., de Lima, D.O., da Silva, H.A., Lent, R.: Automatic isotropic fractionation for large-scale quantitative cell analysis of nervous tissue. J Neurosci Methods **212**(1), 72-8 (2013)
31. Dhikav, V., Anand, K.: Principles and Practice of Neuropsychopharmacology: A Clinical Reference for Residents, Physicians, and Biomedical Scientists. Thieme, New York (2022)

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97554-7_21

32. Hindley, N., Sanchez Avila, A., Henstridge, C.: Bringing synapses into focus: Recent advances in synaptic imaging and mass-spectrometry for studying synaptopathy. Front. Synaptic Neurosci. **15**:1130198. doi: 10.3389/fnsyn.2023.1130198 (2023)
33. DeFelipe, J.: The evolution of the brain, the human nature of cortical circuits, and intellectual creativity. Front Neuroanat. **5**:29. doi: 10.3389/fnana.2011.00029 (2011)
34. Nowinski, W.L.: High performance computing requirements for the calculation of a morphology model of the human whole brain at the nanoscale. Proc. 2023 IEEE International Conference on Bioinformatics and Biomedicine;3321-3327. 10.1109/BIBM58861.2023.10385605 (2023)
35. Peng, Y., Mittermaier, F.X., Planert, H., Schneider, U.C., Alle, H., Geiger, J.R.P.: High-throughput microcircuit analysis of individual human brains through next-generation multineuron patch-clamp. Elife **8**, 1–52. https://doi.org/10.7554/eLife.48178 (2019)
36. Domínguez-Álvaro, M., Montero-Crespo, M., Blazquez-Llorca, L., DeFelipe, J., Alonso-Nanclares, L.: 3D ultrastructural study of synapses in the human entorhinal cortex. Cereb Cortex **31**(1):410–425. https://doi.org/10.1093/cercor/bhaa233 (2021)
37. Cano-Astorga, N., DeFelipe, J., Alonso-Nanclares, L.: Three-dimensional synaptic organization of Layer III of the human temporal neocortex. Cereb Cortex **31**(10), 4742–4764. https://doi.org/10.1093/ cercor/bhab120 (2021)
38. Van Essen, D.C., Smith, S.M., Barch, D.M., Behrens, T.E., Yacoub, E., Ugurbi,l K.: WU-Minn HCP Consortium. The WU-Minn human connectome project: an overview. NeuroImage **80**, 62–79, https://doi. org/10.1016/j.neuroimage.2013.05.041 (2013)
39. Paquola, C., Seidlitz, J., Benkarim, O., Royer, J., Klimes, P., Bethlehem, R.A.I., Larivière, S, Vos de Wael, R., Rodríguez-Cruces, R., Hall, J.A., et al.: A multi-scale cortical wiring space links cellular architecture and functional dynamics in the human brain. PLoS Biol. **18**(11):e3000979. https://doi.org/10.1371/journal. pbio.3000979 (2020)
40. Santuy, A., Tomás-Roca, L., Rodríguez, J.R., González-Soriano, J., Zhu, F., Qiu, Z., Grant, S.G.N., DeFelipe J., Merchan-Perez A.: Estimation of the number of synapses in the hippocampus and brain-wide by volume electron microscopy and genetic labeling. Sci Rep. **10**(1), 14014. doi: 10.1038/s41598-020-70859-5 (2020)
41. Choi, C.Q: The beating heart of the world's first exascale supercomputer. IEEE Spectrum, 24 June 2022. https://spectrum.ieee.org/frontier-exascale-supercomputer (2022)
42. https://www.anl.gov/aurora, last accessed 2024.12.12
43. https://www.ebrains.eu/, last accessed 2024.12.17
44. Hider, R. Jr, Kleissas, D., Gion, T., Xenes, D., Matelsky, J., Pryor, D., Rodriguez, L., Johnson, E.C., Gray-Roncal, W., Wester, B.: The Brain Observatory Storage Service and Database (BossDB): A Cloud-native approach for petascale neuroscience discovery. Front Neuroinform. **16**, 828787. doi: 10.3389/fninf.2022.828787 (2022)
45. Cano-Astorga, N., Plaza-Alonso, S., DeFelipe, J., Alonso-Nanclares, L.: Volume electron microscopy analysis of synapses in primary regions of the human cerebral cortex. Cereb Cortex **34**(8), bhae312. doi: 10.1093/cercor/bhae312 (2024)
46. Chindemi, G., Abdellah, M., Amsalem, O., Benavides-Piccione, R., Delattre, V., Doron, M., Ecker, A., Jaquier, A.T., King, J., Kumbhar, P., et al.: A calcium-based plasticity model for predicting long-term potentiation and depression in the neocortex. Nat Commun. **13**(1), 3038. https://doi.org/10.1038/s41467-022-30214-w (2022)
47. Santuy, A., Rodríguez, J-R., DeFelipe, J., Merchán-Pérez, A.: Volume electron microscopy of the distribution of synapses in the neuropil of the juvenile rat somatosensory cortex. Brain Struct Funct. **223**(1), 77–90. https://doi.org/10.1007/s00429-017-1470-7 (2018)

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97554-7_21