Data for Brain Reference Architecture of NY24CanonicalCorticalMicrocircuitsInference

Canonical cortical microcircuits reference architecture for cognitive inference

Naohiro Yamauchi[a*], Yoshimasa Tawatsuji[bc], Yudai Suzuki[bc], Kenji Doya[a], Hiroshi Yamakawa[bc]

[a]Okinawa Institute of Science and Technology, Okinawa, Japan

[b]The University of Tokyo, Tokyo, Japan

[c]The Whole Brain Architecture Initiative, Tokyo, Japan

[*]Corresponding author: Naohiro Yamauchi; naohiro.yamauchi@oist.jp

Abstract

Canonical cortical microcircuits (CCMs) is the six-layer structure preserved throughout the mammalian neocortex and is thought as a fundamental computational unit. This dataset reverse-engineers CCMs and presents a computational model to achieve cognitive inference. The data consist of the anatomical connectivity of CCMs and the functions hierarchically achieved from each uniform circuit. First, information on the anatomical connections of CCMs was collected from seven review papers. Next, dynamic Bayesian inference was determined as the algorithm for cognitive inference, which CCMs can implement. Finally, we describe how top-level functions are achieved from excitatory neural populations and circuit motif based on inhibitory neural populations, assigning output semantics to each excitatory neural population. The data are described in a brain reference architecture format and stored in the BRA data repository. This dataset provides experimentally testable hypotheses about neural activity patterns in cortical layers.

Keywords: Brain Reference Architecture; Canonical Cortical Microcircuits; Dynamic Bayesian inference

Author roles:

Naohiro Yamauchi Data curation, Methodology, Investigation, Visualization, Writing-original draft

Yoshimasa Tawatsuji Investigation, Validation, Supervision, Writing-review & editing

Yudai Suzuki Investigation, Validation, Writing-review & editing

Kenji Doya Funding acquisition, Supervision, Writing-review & editing

Hiroshi Yamakawa Conceptualization, Funding acquisition, Supervision, Project administration, Writing-review & editing

1 Context

Brain Reference Architecture (BRA) is the reference architecture for software that realizes cognitive and behavioral functions in a brain-like manner. The architecture primarily consists of the mesoscopic-level anatomical data of the brain and the data of one or more functional mechanisms that are consistent with that knowledge(Yamakawa, 2021). BRA consists of Brain Information Flow (BIF), which represents structural knowledge of the brain, and Hypothetical Component Diagram (HCD)/Funciton Realization Graph (FRG), which represent brain functionality.

The canonical cortical microcircuits (CCMs) is a conserved six-layered anatomical structure in the mammalian neocortex(Felleman & Van Essen, 1991; Larkum, 2013). In humans, millions of minicolumns form this structure, functioning as presumed computational units(Maruoka, Kubota, Kurokawa, Tsuruno, & Hosoya, 2011; Mountcastle, 1997). While there are slight variations, such as in the thickness of layer 4, the six-layer structure is preserved across cortical regions involved in sensory processing, motor functions, and higher cognitive functions(Larkum, 2013; Weiler, Wood, Yu, Solla, & Shepherd, 2008). Traditionally, information processing within CCMs was understood as a sequential pathway: from the thalamus to layer 4, then from layer 4 to layers 2/3, and subsequently to layer 5. However, recent studies have revealed a more complex pattern of anatomical and functional connectivity, including direct thalamic projections to layer 5(Audette, Urban-Ciecko, Matsushita, & Barth, 2017; Constantinople & Bruno, 2013). Neuronal populations within CCMs are broadly classified into excitatory and inhibitory cell types. Excitatory cells, which are distinguished by specific projection patterns, are thought

to play a central role in computational processes and include intratelencephalic (IT), pyramidal tract (PT), and corticothalamic (CT) neurons(J. A. Harris et al., 2019; Shepherd & Yamawaki, 2021). Inhibitory cells, which have less specific projection patterns, are thought to support excitatory computations through circuit motifs and include parvalbumin (PV), somatostatin (SST), vasoactive intestinal peptide (VIP), and neurogliaform (NGF) neurons(Tremblay, Lee, & Rudy, 2016; Wang & Yang, 2018).

Table 1: Abbreviations and formal names of the brain regions used in this research.

Abbreviations	Formal names
CCMs	Canonical cortical microcircuits
Cx.High	Higher cortex
Cx.Low	Lower cortex
THM	Thalamus
BN	Basal ganglia
VTA	Ventral tegmental area

Table 2: Abbreviations and formal names of the laminar organization of canonical cortical microcircuits used in this research.

Abbreviations	Formal names
L1	Layer 1
L2	Layer 2
L3	Layer 3
L2_3	Layer 2&3
L4	Layer 4
L5	Layer 5
L6	Layer 6

Table 3: Abbreviations and formal names of the cell types of canonical cortical microcircuits used in this research.

Abbreviations	Formal names
PY	Pyramidal
E	Excitatory
IT	Intratelencephalic
PT	Pyramidal tract
CT	Corticothalamic
IN	Inhibitory
PV	Parvalbumin
SST	Somatostatin
VIP	Vasoactive intestinal peptide
NGF	Neurogliaform

Due to their structural importance and ubiquity, CCMs have been the subject of several hypotheses about their computational functions(Bastos et al., 2012; Doya, 2021; Friston, Parr, & de Vries, 2017; George & Hawkins, 2009; Kermani Nejad, Anastasiades, Hertäg, & Costa, 2024; Miyashita, 2024; Rao, 2024). Theories such as the Bayesian brain hypothesis, Bayesian belief propagation, and predictive coding suggest that neural populations in each layer contribute to cognitive inference—the process by which the brain integrates sensory input and prior knowledge to draw conclusions and make predictions, or derive meaning—by encoding variables through anatomical connections(Bastos et al., 2012; George & Hawkins, 2009; Miyashita, 2024). Recently, models utilizing self-supervised learning to generate predictions have been proposed(Kermani Nejad et al., 2024). Additionally, models aiming to unify inference and decision-making—the cognitive process of selecting a course of action among multiple alternatives based on goals, predictions, and contextual information—within the neocortex have been developed, including those based on active inference, active predictive coding, and the duality of inference and control implemented within CCMs(Doya, 2021; Friston et al., 2017; Rao, 2024).

In this study, we outline how cognitive inference could be achieved through specific excitatory neural populations and circuit motifs involving inhibitory populations within CCMs. We assign output semantics to each neural population in the following way: modulated sensory evidence to layer 2 IT neurons, predictions to layer

3 IT neurons, sensory evidence to layer 4 excitatory neurons, predictive priors to layer 5 IT neurons, posterior predictions to layer 5 PT neurons, and maximum a posteriori (MAP) estimates to layer 6 IT neurons.

This data aims to connect the anatomical structure and functional roles of CCMs by developing a computational model capable of dynamic Bayesian inference (DBI). Through a systematic decomposition of functions necessary for them, using both top-down and bottom-up approaches, we describe the roles of excitatory and inhibitory populations within CCMs. A computational model to achieve decision-making based on the anatomical structure of CCMs has also been developed in a similar way and described in a different publication, with some overlap in data and descriptions.

2 Method

SCID method The series of procedures followed to produce the dataset according to structure-constrained interface decomposition (SCID) method (Yamakawa, 2021).

The brief introduction of three steps of SCID method is given as follows:

Step 1. Brain Information Flow (BIF) registering and provisonary creation of Hypothetical Component Diagram. This steps include (a) surveying anatomical knowledge in specific brain region (ROI: region of interest), (b) following determination of ROI and TLF (top-level function) consistently and (c) creation of a provisionary component diagram (called HCD)

Step 2. Enumerating candidate component diagram.

Step 3. Rejecting diagram that are inconsistent with scientific knowledge.

You can see more details about these steps in (Yamakawa, 2021).

Motif definition A motif is a frequently occurring pattern in neural circuits and represents a fundamental functional unit of complex neural networks(Braganza & Beck, 2018). Motifs consist of nodes, representing neurons, and edges, representing the connections between neurons. In this study, we focus on motifs involving inhibitory neurons, examining their contributions to circuit-implemented algorithms(Tremblay et al., 2016). Figure 1 illustrates the motifs employed in this research. Here, C denotes the capabilities achieved by the motif; for example, in the left panel of Figure 1, feedforward inhibition enables coincidence detection. X indicates input, Z represents nodes, black edges denote excitatory connections, and blue edges denote inhibitory connections.

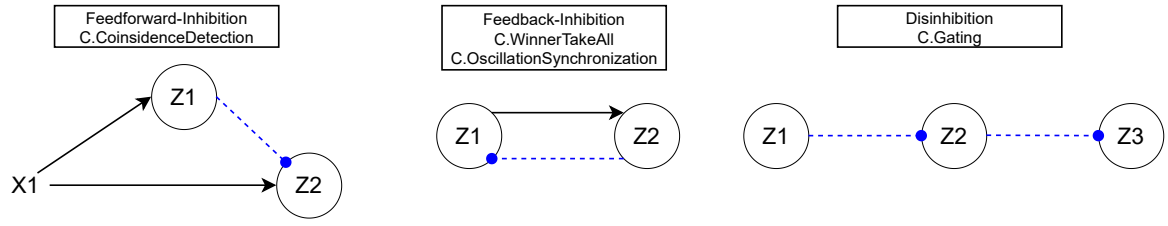

Figure 1: Circuit motifs and corresponding capabilities used in this study.

Dynamic Bayesian inference Dynamic Bayesian inference is a probabilistic framework for updating predictions about a internal state that evolves over time, integrating prior knowledge with incoming noisy sensory data to estimate latent variables(Bishop, 2006). This approach relies on probabilistic graphical models, such as Hidden Markov Models (HMMs). Assuming that the state transition model and observation model are pretrained, we hypothesize that this graphical model can be implemented in canonical cortical microcircuits, mapping its components to cortical layers(Bastos et al., 2012). Specifically, sensory input from the thalamus is converted into likelihood information by the observation model intrinsic to layer 4, then transmitted to layer 2. Simultaneously, the posterior distribution from the previous time step is input as a prior, combined with state predictions derived from the state transition model in layer 3, and transformed into a predictive prior by layer 5 IT neurons. The likelihood and predictive prior are then multiplied and normalized by inhibitory neurons, forming the posterior prediction in layer 5 PT neurons.

Sampling strategy The dataset was constructed by gathering and integrating data from seven publications(Billeh et al., 2020; Dura-Bernal et al., 2023; K. D. Harris & Mrsic-Flogel, 2013; Shipp, 2007; Thomson, 2007; Tremblay et al., 2016; Vitrac & Benoit-Marand, 2017), including reviews and modeling papers, authored by researchers involved in cortical experiments and theoretical neuroscience. The selection of references considered multiple factors, including the journal of publication, the comprehensiveness of anatomical descriptions, the level of detail specific to particular circuits, and consistency with other literature. Detailed information on the referenced publications, including titles, authors, journals, and publication years, is available in the "References"

sheet of the dataset. The motifs used to construct the FRG data were chosen based on prior empirical studies of cortical neural circuits and functional requirements essential for algorithm implementation.

3 Dataset Description

Repository location BRA Editorial System (BRAES) https://sites.google.com/wba-initiative.org/braes/data

Object name and versions Please refer to the "Project" sheet in the BRA data for the more detail of data summary.

Table 4: BRA DATA SUMMARY

BRA Data			
Object Name	Template	Including Content(s)	
		BIF	HCD/FRG
NY24CanonicalCorticalMicrocircuitsInference.bra	*version 2.1.1*	√	√

Table 5: BRA IMAGE SUMMARY

Graphic Files: BIF Image, HCD Image, FRG Image	
File Type	Object Name
BIF Image	NY24CanonicalCorticalMicrocircuitsInferenceBIF.xml
HCD Image	NY24CanonicalCorticalMicrocircuitsInferenceHCD.xml
FRG Image	NY24CanonicalCorticalMicrocircuitsInferenceFRG.xml

Creation dates 2024-8-5 to 2025-1-18.

Language English.

License The open license under which the data has been deposited (CC-BY 4.0).

Publication date 2025-1-18.

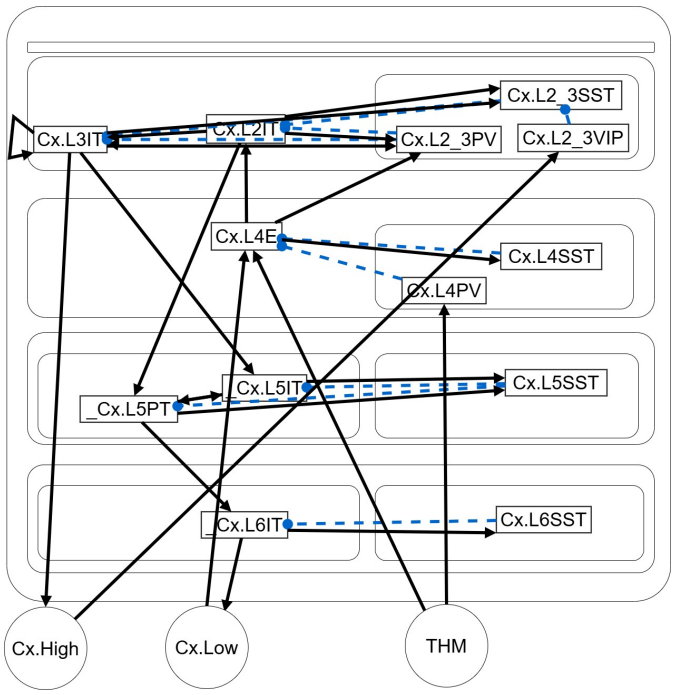

Figure 2: NY24CanonicalCorticalMicrocircuitsInferenceHCD.

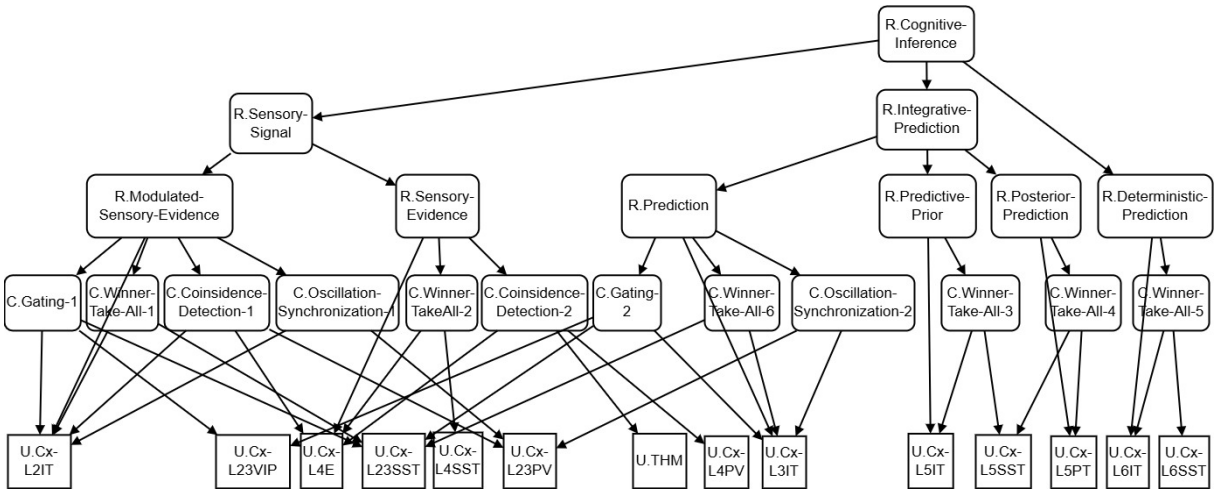

Figure 3: NY24CanonicalCorticalMicrocircuitsInferenceFRG.

4 Caveats for Data Usage

The brain-referenced architecture (BRA) data encompasses BIF (Brain Information Flow), HCD (Hierarchical Component Diagram), and FRG (Functional Realization Graph) data, offering potential reuse for researchers in neuroscience and information science. This dataset provides standardized descriptions across various brain regions, facilitating integration and comparison of distinct neural areas. Beyond contributing to a deeper understanding of brain anatomy and function, these data can support the construction of simulation models.

This study includes several considerations that users should note. Notably, it incorporates inhibitory neuron subtypes and their connections into the anatomical structure of canonical cortical microcircuits. However, due to technical limitations in current experimental methodologies, projections to inhibitory neurons are underrepresented in the literature. Thus, the current anatomical data on canonical cortical microcircuits remains incomplete. The constructed computational component diagram (HCD) aligns with anatomical connections documented in BIF, utilizing only verified connections. However, it does not account for a range of possible projections, as selections were made to fulfill computational requirements based on a subset of BIF connections. Although the canonical cortical microcircuit serves as the fundamental unit of various cortical functions, the present HCD is limited to functions related to inference and decision-making (described in a separate paper, with some overlap in data and descriptions). Future work is required to extend the HCD to support other cortical functions.

Funding Statement

This research was supported by the Matsuo-Iwasawa lab and the ANRI fellowship.

Competing interests

Yoshimasa Tawatsuji and Hiroshi Yamakawa are BRAES managers but did not participate in the editorial process or decisions related to this manuscript.

References

Audette, N. J., Urban-Ciecko, J., Matsushita, M., & Barth, A. L. (2017). Pom thalamocortical input drives layer-specific microcircuits in somatosensory cortex. *Cerebral Cortex*, *28*(4), 1312–1328. DOI: 10.1093/cercor/bhx044

Bastos, A. M., Usrey, W. M., Adams, R. A., Mangun, G. R., Fries, P., & Friston, K. J. (2012). Canonical microcircuits for predictive coding. *Neuron*, *76*(4), 695–711.

Billeh, Y. N., Cai, B., Gratiy, S. L., Dai, K., Iyer, R., Gouwens, N. W., ... Arkhipov, A. (2020). Systematic integration of structural and functional data into multi-scale models of mouse primary visual cortex. *Neuron*, *106*(3), 388–403.e18. DOI: 10.1016/j.neuron.2020.01.040

Bishop, C. (2006). *Pattern recognition and machine learning.* Springer.

Braganza, O., & Beck, H. (2018). The circuit motif as a conceptual tool for multilevel neuroscience. *Trends in Neurosciences*, *41*(3), 128–136. DOI: 10.1016/j.tins.2018.01.002

Constantinople, C. M., & Bruno, R. M. (2013). Deep cortical layers are activated directly by thalamus. *Science*, *340*(6140), 1591–1594. DOI: 10.1126/science.1236425

Doya, K. (2021). Canonical cortical circuits and the duality of bayesian inference and optimal control. *Current Opinion in Behavioral Sciences*, *41*, 160–167. DOI: 10.1016/j.cobeha.2021.07.003

Dura-Bernal, S., Neymotin, S. A., Suter, B. A., Dacre, J., Moreira, J. V., Urdapilleta, E., ... Lytton, W. W. (2023). Multiscale model of primary motor cortex circuits predicts in vivo cell-type-specific, behavioral state-dependent dynamics. *Cell Reports*, *42*(6), 112574. DOI: 10.1016/j.celrep.2023.112574

Felleman, D. J., & Van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex*, *1*(1), 1-47. DOI: 10.1093/cercor/1.1.1-a

Friston, K. J., Parr, T., & de Vries, B. (2017). The graphical brain: Belief propagation and active inference. *Network Neuroscience*, *1*(4), 381–414. DOI: 10.1162/netn$_{a0}$0018

George, D., & Hawkins, J. (2009). Towards a mathematical theory of cortical micro-circuits. *PLoS Computational Biology*, *5*(10), e1000532. DOI: 10.1371/journal.pcbi.1000532

Harris, J. A., Mihalas, S., Hirokawa, K. E., Whitesell, J. D., Choi, H., Bernard, A., ... Zeng, H. (2019). Hierarchical organization of cortical and thalamic connectivity. *Nature*, *575*(7781), 195–202. DOI: 10.1038/s41586-019-1716-z

Harris, K. D., & Mrsic-Flogel, T. D. (2013). Cortical connectivity and sensory coding. *Nature*, *503*(7474), 51–58. DOI: 10.1038/nature12654

Kermani Nejad, K., Anastasiades, P., Hertäg, L., & Costa, R. P. (2024). Self-supervised predictive learning accounts for cortical layer-specificity. DOI: 10.1101/2024.04.24.590916

Larkum, M. (2013). A cellular mechanism for cortical associations: an organizing principle for the cerebral cortex. *Trends in Neurosciences*, *36*(3), 141–151. DOI: 10.1016/j.tins.2012.11.006

Maruoka, H., Kubota, K., Kurokawa, R., Tsuruno, S., & Hosoya, T. (2011). Periodic organization of a major subtype of pyramidal neurons in neocortical layer v. *The Journal of Neuroscience*, *31*(50), 18522–18542. DOI: 10.1523/jneurosci.3117-11.2011

Miyashita, Y. (2024). Cortical layer-dependent signaling in cognition: Three computational modes of the canonical circuit. *Annual Review of Neuroscience*, *47*(1), 211–234. DOI: 10.1146/annurev-neuro-081623-091311

Mountcastle, V. (1997). The columnar organization of the neocortex. *Brain*, *120*(4), 701–722. DOI: 10.1093/brain/120.4.701

Rao, R. P. (2024). A sensory–motor theory of the neocortex. *Nature Neuroscience*, *27*(7), 1221–1235.

Shepherd, G. M. G., & Yamawaki, N. (2021). Untangling the cortico-thalamo-cortical loop: cellular pieces of a knotty circuit puzzle. *Nature Reviews Neuroscience*, *22*(7), 389–406. DOI: 10.1038/s41583-021-00459-3

Shipp, S. (2007). Structure and function of the cerebral cortex. *Current Biology*, *17*(12), R443–R449.

Thomson, A. M. (2007). Functional maps of neocortical local circuitry. *Frontiers in Neuroscience*, *1*(1), 19–42. DOI: 10.3389/neuro.01.1.1.002.2007

Tremblay, R., Lee, S., & Rudy, B. (2016). Gabaergic interneurons in the neocortex: From cellular properties to circuits. *Neuron*, *91*(2), 260–292. DOI: 10.1016/j.neuron.2016.06.033

Vitrac, C., & Benoit-Marand, M. (2017). Monoaminergic modulation of motor cortex function. *Frontiers in Neural Circuits*, *11*. DOI: 10.3389/fncir.2017.00072

Wang, X.-J., & Yang, G. R. (2018). A disinhibitory circuit motif and flexible information routing in the brain. *Current Opinion in Neurobiology*, *49*, 75–83. DOI: 10.1016/j.conb.2018.01.002

Weiler, N., Wood, L., Yu, J., Solla, S. A., & Shepherd, G. M. G. (2008). Top-down laminar organization of the excitatory network in motor cortex. *Nature Neuroscience*, *11*(3), 360–366. DOI: 10.1038/nn2049

Yamakawa, H. (2021). The whole brain architecture approach: Accelerating the development of artificial general intelligence by referring to the brain. *Neural Networks*, *144*, 478–495. DOI: 10.1016/j.neunet.2021.09.004