2021 Special Issue on AI and Brain Science: Brain-inspired AI

Hippocampal formation-inspired probabilistic generative model

Akira Taniguchi [a,*], Ayako Fukawa [b], Hiroshi Yamakawa [b,c,d]

[a] *Ritsumeikan Univervsity, 1-1-1 Noji-Higashi, Kusatsu, Shiga 525-8577, Japan*
[b] *The Whole Brain Architecture Initiative, Nishikoiwa 2-19-21, Edogawa-ku, Tokyo, 133-0057, Japan*
[c] *The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan*
[d] *RIKEN, 6-2-3, Furuedai, Suita, Osaka 565-0874, Japan*

ARTICLE INFO

ABSTRACT

In building artificial intelligence (AI) agents, referring to how brains function in real environments can accelerate development by reducing the design space. In this study, we propose a probabilistic generative model (PGM) for navigation in uncertain environments by integrating the neuroscientific knowledge of hippocampal formation (HF) and the engineering knowledge in robotics and AI, namely, simultaneous localization and mapping (SLAM). We follow the approach of brain reference architecture (BRA) (Yamakawa, 2021) to compose the PGM and outline how to verify the model. To this end, we survey and discuss the relationship between the HF findings and SLAM models. The proposed *hippocampal formation-inspired probabilistic generative model* (HF-PGM) is designed to be highly consistent with the anatomical structure and functions of the HF. By referencing the brain, we elaborate on the importance of integration of egocentric/allocentric information from the entorhinal cortex to the hippocampus and the use of discrete-event queues.

© 2022 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

1. Introduction

In building artificial intelligence (AI) agents, referring to how brains function in real environments can accelerate development by reducing the design space. Hippocampal formation (HF) supports crucial neural capabilities, such as spatial cognition, self-localization for navigation, mapping, and episodic memory. In neuroscience, HF and its functions have attracted increasing attention in recent years. The hippocampus has long been considered the brain region responsible for configuring the cognitive map (O'keefe & Nadel, 1978; Tolman, 1948). To this end, designated neurons, such as place cells in the hippocampus (O'keefe & Nadel, 1978) and grid cells in the medial entorhinal cortex (MEC), exist to execute these functions (Fyhn et al., 2004; Hafting et al., 2005). From the perspective of computational neuroscience, numerous computational model-based studies have focused on functions involving the hippocampus (Banino et al., 2018; Kowadlo et al., 2019; Madl et al., 2015; Milford et al., 2004; Schapiro et al., 2017; Scleidorovich et al., 2020). Alongside these computational studies, the use of brain-inspired AI and intelligent robotics is crucial to the implementation of these spatial functions. From an engineering perspective, simultaneous localization and mapping (SLAM) (Thrun et al., 2005) represents a typical approach in computational geometry and robotics. Spatial cognition and place understanding are important challenges that must be overcome to facilitate the advance of robotics (Taniguchi et al., 2019). However, despite the abundancy of neuroscience knowledge related to HF and the progress in AI technology, combining knowledge from both fields and applying it to robotics remains a major challenge.

Purposes: This study aims to bridge the gap between neuroanatomical/biological findings of the HF and engineering technologies of probabilistic generative models (PGMs), particularly in AI and robotics. This paper is a feasibility study on the methodology proposed by Yamakawa (2021). We establish a correspondence between the function/structure of the HF in neuroscience and spatial cognitive methods in robotics. The main objectives of this study are as follows.

- To provide suggestions for the construction of a computational model with functions of HF by surveying the association between SLAM in robotics and HF in neuroscience.
- To construct a *brain reference architecture* (BRA) that operates with biologically valid and consistent functions, as a specification for implementing a brain-inspired model.

Type of paper: This report is a hypothesis-suggestion paper that presents a novel argument, interpretation, or model intended to introduce a hypothesis/theory, based on a literature review, and provides the direction for its verification. We

---

* Corresponding author.
E-mail addresses: a.taniguchi@em.ci.ritsumei.ac.jp (A. Taniguchi), fukawa@yairilab.net (A. Fukawa), ymkw@wba-initiative.org (H. Yamakawa).

construct a *hippocampal formation-inspired probabilistic generative model* (HF-PGM) as a highly adequate hypothesis by referencing neuroanatomical/biological findings. The adequacy of the constructed model can be evaluated by adopting the *structure-constrained interface decomposition* (SCID) method (Yamakawa, 2021) for hypothesis generation. A *generation-inference process allocation* task is used to solve particular problems related to the mapping of PGMs to brain circuits. This task involves allocating all anatomical connections to either the generative or the inference process. Here, we learn the connections between modules from the brain and utilize engineering technologies for the parts having insufficient findings of the brain structure and function. The available findings on HF have been primarily gathered from studies on rodents and only partially from the human brain. Engineering technologies refer to methods related to spatial awareness (e.g., SLAM, navigation, place recognition, spatial concept formation, and semantic mapping) (Kostavelis & Gasteratos, 2015; Thrun et al., 2005). Thus, we present computational HF models to support a feasible hypothesis from an engineering perspective. Furthermore, we provide suggestions for the construction of methods to study spatial cognitive functions using robotics and AI technology from the perspective of neuroscience. The proposed PGM is designed using an extension of the development method proposed by Yamakawa (2021). Therefore, the adequacy of the architecture is based on the evaluation criteria proposed in the same work. These evaluation criteria require (a) the brain information flow (BIF) to be consistent with the anatomical findings, (b) the hypothetical component diagram (HCD) to be consistent with the structure of the BIF, and (c) the HCD to be able to achieve the expected computational functions. The proposed PGM corresponding to the neural circuitry of the HF is obtained through a design that is a natural extension of SLAM based on engineering practice, and the design procedure is detailed in Section 6.3. Hence, the feasibility (c) of HCD is computationally supported without any new implementations or simulations.

Contributions: The main contributions of this study are as follows.

- We proffer and apply the generation-inference process allocation, an approach that allows neural circuits to be interpreted as PGM, for the first time.
- We clarify that the function and structure of HF can be consistently represented as an extension of the previously reported SLAM models by considering relevant findings on HF and SLAM.
- We show that the phase precession of brain activity in the HF can be formulated as a *discrete-event queue* by repeatedly performing sequential Bayesian inference on the joint probability distribution of the variables.
- We illuminate the direction of future challenges that are important for the development of HF-inspired models through discussions on spatial movement, hierarchies, and physical constraints.

This paper is structured as follows. Section 2 describes the background and motivation. Section 3 describes the methodology for constructing a brain reference architecture and PGM. Section 4 summarizes the neuroscientific findings on HF. Section 5 describes relevant topics, including brain-inspired SLAM, computational models of the HF, concept formation and semantic understanding of location and space, and deep generative models. Section 6 describes the proposed HF-PGM from the perspective of the model structure. We construct a novel PGM that is highly consistently with the neuroanatomical/biological findings of HF by integrating allocentric/egocentric visual information. Section 7 describes the formulation of a discrete-event queue based on the phase precession queue assumption in HF as one of the inference algorithms for PGMs. Section 8 summarizes the findings of this study and discusses future directions.

2. Backgrounds and motivation

2.1. Bridging the gap between two fields

Models facilitated by engineering approaches, such as robotics and AI, are capable of supporting HF modeling. Numerous aspects of the hippocampus remain unknown. Hence, there are limitations to creating an HF model solely based on existing hippocampus-related knowledge in neuroscience and biology. Models for problem-solving based on engineering do not encounter these types of restrictions. It is possible to build a model with functions similar to those of the hippocampus without actively utilizing its knowledge. Such engineering models constructed for solving actual tasks may have implications for neuroscience and biology.

Furthermore, the knowledge on neuroscience is useful for engineering applications in robotics and AI. Robotics generally depends on a particular task and involves practical applications (e.g., accurate 3-dimensional (3D) modeling of the environment. Brain-inspired AI emphasizes learning from the brain and mapping brain circuitry and functions, followed by searching for possible practical applications. Building a system that refers to the brain that actually operates with functions for various tasks opens engineering possibilities from novel perspectives.

Brain reference architecture, which links neuroscience and robotics while enabling their application in robots, is currently an important field of research.

2.2. PGM-based cognitive architecture

Probabilistic models that have succeeded as neocortical computational models are also useful for modeling HF. In particular, PGMs represent the process of generating observational stimuli by assuming dependencies between random variables. Probabilistic inference (e.g., Bayesian) may be employed by adopting PGMs, which infer states behind sensory stimuli as latent variables, enabling the construction of internal representations. The observed variables correspond to the stimuli. Adopting PGMs is valid from the perspective of the Bayesian brain hypothesis, which states, "the brain represents sensory signals probabilistically in the form of probability distributions" (Doya et al., 2007; Knill & Pouget, 2004). Furthermore, a reason for modeling brain-inspired architecture using PGMs is that the Neuro-symbol emergence in the robotics tool kit (SER-KET) architecture can be used both theoretically and practically (Taniguchi et al., 2020). Neuro-SERKET enables PGMs having several functions to have distributed development, which can be used to develop integration modules. This architecture enables the integration of models that imitate other regions in the brain into a whole-brain model. Therefore, the PGMs for HF proposed in this study may be used as a module to represent the whole-brain's integrated cognitive architecture (Taniguchi et al., 2022).

3. Brain reference architecture (BRA) construction

This section describes the methodology for constructing the PGM as a BRA. Fig. 1 shows the relationship of components for the modeling processes. Section 3.1 introduces an overview on BRA. Section 3.2 describes construction process and adequacy evaluation for BIF and HCD. Section 3.3 describes a problem-solving approach for associating anatomical structures in the brain with PGMs (i.e., generation-inference process allocation). The generation-inference process allocation is devised and demonstrated for the first time in this study.
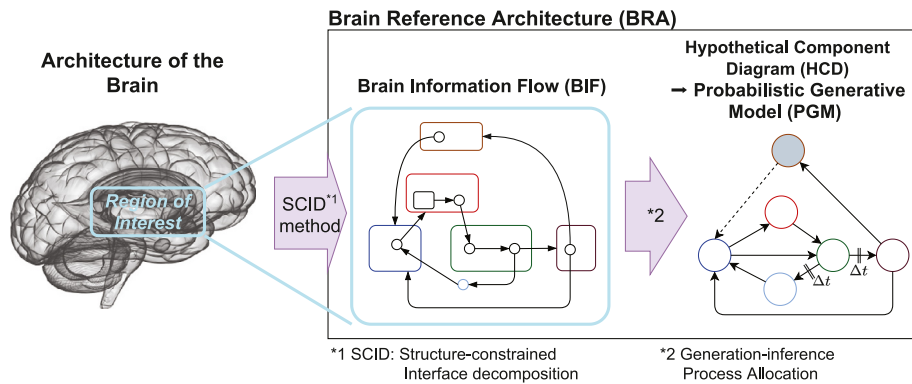
Fig. 1. Overview of modeling processes for brain reference architecture (BRA). The brain information flow (BIF) and hypothetical component diagrams (HCD) are constructed using the structure-constrained interface decomposition (SCID) method. The probabilistic generative model (PGM) is one of the expression formats of the hypothetical component diagrams constructed by generation-inference process allocation.

3.1. Brain reference architecture

This study adopts an approach for building parts of modules in a whole-brain architecture, which is a brain-inspired artificial general intelligence development approach that emphasizes the architecture. Design methodologies for combining machine learning that imitates the mesoscopic-level brain structures of rodents and humans are being studied to gather knowledge on the architecture of the brain. Specifically, the whole-brain architecture initiative standardizes information corresponding to the fulfillment of requirement specifications for brain-inspired software as BRA data[1] and promotes role sharing in the BRA design and utilization. The whole-brain architecture initiative includes a manual[2] for preparing brain reference architecture (BRA) as the standard notation for guiding the development of brain-inspired software.

BRA, which is the reference architecture at the mesoscopic level, represents a consistent description of (i) the brain information flow (BIF) related to anatomical structures and (ii) the hypothetical component diagram (HCD) related to its computational functions (Yamakawa, 2021). BIF is a directed graph comprising partial circuits and connections that represent the anatomical structure of neural circuits in the brain. HCD is a directed graph describing component dependencies, which can be associated with any BIF sub-graph. The functional mechanisms described in HCD are hypothetical, as the name implies, and different hypotheses may be presented by neuroscientists from different perspectives. In many cases, brain-inspired software is implemented by engineers who do not necessarily have a deep understanding of neuroscience. Therefore, it is useful to examine HCD candidates that cannot be clearly dismissed based on current neuroscience knowledge, even if they cannot be confirmed as ground truth.

3.2. Construction process and adequacy evaluation

In this study, we use a *structure-constrained interface decomposition* (SCID) method developed by the whole-brain architecture initiative (Yamakawa, 2021) to design an HCD consistent with the anatomical structure of the HF. Using the SCID method, Fukawa et al. (2020) applied the concept to the neural circuit of HF for the first time and identified a circuit that performs a path integration function on MEC.

We created one plausible HCD by applying steps 1 and 2 (see Appendix A for the detail) mainly to the HF. In this study, PGM is positioned as a form of expression for HCD. Fig. 2 shows a table of PGM modeling processes in the SCID method. The execution result of step 1-A is mainly explained in Section 4. BIF is shown in Fig. 3. The execution result of step 1-B is explained in Section 5. The execution result of step 1-C is explained in Section 6.1. The execution result of Step 2 is explained in Section 6.3. The contents partially examined in Step 3 are described in Section 6.4.

The authenticity of BIF was established by Yamakawa (2021) in the section "Adequacy evaluation of BIF". We further elaborate on this adequacy in Section 4. Similarly, the consistency of HCD with BIF has been described in the section "Adequacy evaluation of HCD" in Yamakawa (2021). The functionality of HCD is ensured by the presence of SLAM models (details are presented in Section 6.3. The model was created to be consistent with BIF, thus ensuring structure-consistency (discussed in Section 6.4).

3.3. Generation-inference process allocation

We describe the task developed in this study to solve particular problems for mapping brain circuits when modeling PGMs. PGMs have the restriction that they must be directed acyclic graphs; hence, loops cannot be represented. (See Appendix B.1 for a basic description of PGMs.) However, some brain circuits have loop structures. In most cases, it is difficult to assign acyclic PGMs to a brain circuit. Furthermore, in ordinary PGMs, signal propagation in the inference process causes signals to propagate in the direction opposite to that of the links used during the generative process. In contrast, in brain neural circuits, signal propagations between regions propagated by electrical spikes that propagate terminally on axons are essentially unidirectional. Generally, modeling a PGM of any existing interarea connection in the brain is a difficult task.

To eliminate this restriction, we adopt a combination of generative/inference models.[3] In other words, we assume that an *amortized variational inference* is introduced, as described in Appendix B.1. The amortized variational inference can define the link structure of the inference process without depending on the link structure of the generative process. The model used for amortized variational inference can be designed with a high degree of freedom, as long as it is consistent with the link structure of the generative process. Even if loops are present, there is a time delay in the signal transduction in the neural circuits of the brain.

---

[1] https://wba-initiative.org/wiki/en/brain_reference_architecture.

[2] BRA Data Preparation Manual: https://docs.google.com/document/d/1_t3W_dkFmGjfBhz3_EEZ2FCYyzrJi_1ZOtoPlOru8dc/edit#heading=h.6766fia4kgtf.

[3] As other solutions, we can use probabilistic models based on factor/undirected graphs instead of generative models.

Modeling process				Described section
PGM modeling	SCID method	Step 1	Step 1-A: Surveying anatomical knowledge in region-of-interest	4. Neuroscientific findings of hippocampal formation
			Step 1-B: Creating a provisory component diagram	5. Relevant topics and correspondence between HF and SLAM
			Step 1-C: Determining region-of-interest and top-level function consistently	6.1. Region-of-interest, top-level function, and input/output
		Step 2: Enumerating candidate component diagrams		6.3. Models and operating principle
		Step 3: Rejecting diagrams that are inconsistent with scientific knowledge		6.4. Consistency of model with scientific knowledge
	Generation-inference process allocation			6.2. Executing generation-inference process allocation

Fig. 2. Modeling processes and section structure by structure-constrained interface decomposition (SCID) method and generation-inference process allocation.

Therefore, we introduce *next-time generation*,[4] which expresses the progress of time for the generation process, as shown in Fig. B.2(b). Therefore, it becomes easy to relate the link structure of PGMs to the structure of actual brain neural circuits.

This allocating task is the generation-inference process allocation. This task consistently allocates all links into a generative or inference model in a dependency network. In neural circuits adjacent to the cortex, attention must be paid to avoid inconsistencies at the cortex interface. If the neuroscientific findings are uncertain, a reasonable and feasible engineering model is chosen. As a prerequisite for performing this task, each node of these variables[5] is principally associated with a uniform circuit (see Yamakawa (2021)), which is the minimum descriptive unit of the BIF.

Generally, applying generation-inference process allocation to any existing interarea connections in the brain is not an easy task. However, the major interarea connections of the neocortex can be allocated to either the generative or inference process. In the neocortex, a feedforward pathway transmits signals from lower to higher areas while processing signals received by sensors, and a feedback pathway transmits signals in the opposite direction (Markov et al., 2013, 2014). In computational neuroscience theories (e.g., Bayesian brain Doya et al. (2007), Knill and Pouget (2004) and predictive coding (Rao & Ballard, 1999)), inference and generation are assumed to be processed by the feedforward and feedback pathways, respectively. Therefore, the flow of the inference model, represented by the dotted arrow in the graphical model, is associated with the pathway of the feedforward signal, and the flow of the generative model represented by the solid arrow is associated with the pathway of the feedback signal.

4. Neuroscientific findings of hippocampal formation

In this section, we survey the connections and functions of HF and its surrounding areas. The core region of interest of the HF focused in this study includes the cornu ammonis-1 and −3 areas (CA1 and CA3) and dentate gyrus (DG) in the hippocampus, lateral/medial entorhinal cortex (LEC/MEC), subiculum (Sb), and parasubiculum (ParaSb). The presubiculum (PreSb), perirhinal cortex (PER), postrhinal cortex (POR), retrosplenial cortex (RSC), and medial septum are also investigated as areas adjacent to the HF, that is, as regions for sending and receiving the signal.

4.1. Anatomical and physiological findings

With respect to neuroscientific findings, we conducted a survey by adding detailed findings on LEC and hippocampus to the findings in the review of Fukawa et al. (2020) centered on the hippocampus and MEC. These findings are also based on an original anatomical review of the structure of HF (Amaral & Witter, 1989; Witter et al., 2000). Fig. 3 shows the connection relationship of the HF circuit, which includes the hippocampus, Sb, PreSb, ParaSb, and entorhinal cortex. The hippocampus comprises DG, CA1, and CA3. The entorhinal cortex is divided into MEC and LEC. CA1 and Sb are further separated distally and proximally (Knierim et al., 2014). In this study, MEC is divided into six layers: I, II, III, Va, Vb, and VI, and LEC is similarly divided into six layers: I, II, III, IV, V, and VI (Fukawa et al., 2020; Shepherd & Grillner, 2013). MEC IV is excluded because it has few neurons. As input to LEC, a connection exists from the POR and PER to LEC II and III (Nilssen et al., 2019). The output from LEC has a connection from LEC II to DG and CA3 and from LEC III to CA1 and Sb. From CA3, there are two connections, Sb proximal through CA1 distal and Sb distal through CA1 proximal (Knierim et al., 2014). The nonspatial signal is conveyed from LEC to the hippocampus (Hargreaves et al., 2005). The MEC and LEC signals are integrated by the DG and CA3 in the hippocampus (Chen et al., 2013; Knierim et al., 2014).

Neural cells with various expressions have been observed in the hippocampus and its surrounding areas. Place cells exist in CA3 and CA1 (O'keefe & Nadel, 1978); they are active only when an animal enters a specific place in the environment; they do not fire elsewhere. It is thought that an environmental cognitive map is stored as a neural circuit according to the place cells (O'keefe & Nadel, 1978). Grid cells exist in MEC, ParaSb, and PreSb (Hafting et al., 2005). Although the activity of place cells represents one place in the environment, grid cells are activated in a hexagonal grid at multiple places in the environment. Head-direction cells exist in MEC, PreSb, ParaSb, and RSC (Grieves & Jeffery, 2017; Taube, 2007; Taube et al., 1990); they are active only when the head faces a specific direction, regardless of the location of the viewer. Border cells and boundary vector cells are cells in Sb and MEC that selectively fire near the border, irrespective of whether there are objects or obstacles (Lever et al., 2009). Speed cells exist in MEC and exhibit activity that depends on the speed of movement of the subject (Hinman et al., 2016; Kropff et al., 2015). Spatial view cells that respond to the current landscape are known to exist on the primate's hippocampus (Rolls, 2013). Furthermore, event cells have been discovered that represented the content and order of the events experienced (Terada et al., 2017). Events that are not evenly spaced in time are processed centrally by the LEC and stored only when the state changes.

---

[4] It means assuming state–space models for time-series data.

[5] Nodes are basically random variables; furthermore, they can be represented as temporary variables during the calculation process of deterministic variables.
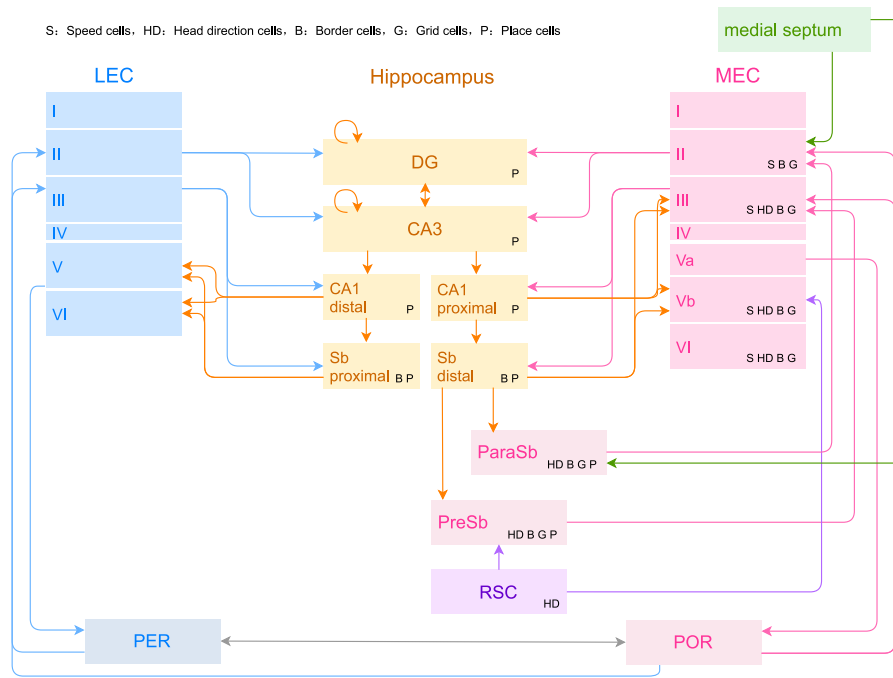
S: Speed cells, HD: Head direction cells, B: Border cells, G: Grid cells, P: Place cells

Fig. 3. Brain information flow (BIF) represents the connection of neural circuits between HF and its surrounding areas. It is based on anatomical studies and reviews (Amaral & Witter, 1989; Fukawa et al., 2020; Witter et al., 2000). Blue, lateral entorhinal cortex (LEC) and perirhinal cortex (PER); red, medial entorhinal cortex (MEC), presubiculum (PreSb), parasubiculum (ParaSb) and postrhinal cortex (POR); yellow, the hippocampus and subiculum (Sb); green, the medial septum; and purple, retrosplenial cortex (RSC). The hippocampus includes the cornu ammonis-1 and -3 areas (CA1 and CA3) and dentate gyrus (DG). Colored arrows indicate that the signal is sent from the area indicated by that color to the other area at the tip of the arrow. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

4.2. Functions

HF is a brain region that controls short-term and episodic memory in vertebrates. It is deeply involved in spatial memory functions, such as spatial learning and exploration. The behavior of an animal when traveling to a destination by selecting an appropriate route while simultaneously acquiring environmental information is called navigation. Cognitive mapping, where animals form a map of the spatial positional relationship of various items in the environment by exploration and act accordingly using this map, is proposed as a psychological concept deeply related to navigation (Giocomo et al., 2011; Tolman, 1948). The hippocampus is thought to be crucial to the formation of cognitive maps in mammals.

Self-localization, where animals consistently recognize their current position, is indispensable for the navigation and formation of cognitive maps. One of the functions that support the self-localization ability of animals is path integration (McNaughton et al., 2006). Path integration is a function that outputs self-position after movement upon input of the initial position, head-direction, and movement signals, including speed and movement direction (Raudies et al., 2015). These researchers argued that the region responsible for path integration was either MEC II stellate cells or MEC III. Path integration is calculated using grid cells (Gil et al., 2018). Their firing pattern represents metric properties, and the firing pattern of place cells represents the position index information without metric properties (Buzsáki & Moser, 2013; Fukawa et al., 2020). The prospective speed signal is calculated by MEC III or MEC Vb. Furthermore, in the hippocampus, there is a loop structure wherein the movement signal by path integration in MEC and signal from the LEC are integrated to obtain accurate position information. DG and CA3 execute the functions of pattern
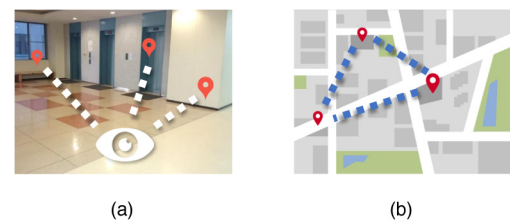

Fig. 4. Egocentric and allocentric representations of space. (a) Egocentric visual information. First-person perspective and self-centered coordinate system. (b) Allocentric visual information. Objective perspective and object center coordinate system.

separation and pattern completion, respectively (Bakker et al., 2008). Pattern separation is the ability to identify the difference between two perceptual patterns, and pattern completion is the ability to generalize and complement using similar signals from partial observation and noisy environments.

LECs and MECs have been reported to perform different processing functions owing to the difference in the signals sent from PER and POR. LEC and MEC process proximal and distal landmarks, respectively (Kuruvilla et al., 2020). In addition, LEC processes signals from an egocentric perspective, whereas MEC processes signals from an allocentric perspective (Alexander et al., 2020; Bicanski & Burgess, 2018; Byrne et al., 2007; Deshmukh & Knierim, 2011; Hargreaves et al., 2005; Wang et al., 2020). Fig. 4 shows the difference between egocentric and allocentric perspectives. Considering these factors, two types of cognitive maps are generated: survey and route maps (Shemyakin, 1962). Therefore, LEC and MEC are expected to be responsible for spatial cognition of the route and survey maps, respectively. These two

types of views are used in car navigation systems and online map apps (e.g., Google Street View).

5. Relevant topics and correspondence between HF and SLAM

We introduce the research on computational models related to spatial cognition methods, including SLAM. Furthermore, we discuss the association between functions of HF and computational models based on the content covered in Appendix B and Section 4. First, we introduce the HF computational models and brain-inspired SLAM in Section 5.1. Second, we introduce the models relating to place category formation and investigate their relationship with HF in Section 5.2. Finally, the neural-network models and their associations with PGMs are described in Section 5.3.

5.1. Computational models for HF and brain-inspired SLAM

Herein, we describe the HF computational model and the relationship between HF and SLAM. Several studies have discussed the relationship between SLAM, partially observable Markov decision process graphical models, their probabilistic inferences, and HF circuits and functions (Bermudez-Contreras et al., 2020; Gaussier et al., 2019; Madl et al., 2018; Penny et al., 2013). SLAM functions mainly correspond to the loop circuits of the MEC and hippocampus (Fukawa et al., 2020). Bermudez-Contreras et al. (2020) discussed the association with artificial neural networks and reinforcement learning in spatial navigation. Furthermore, several studies were conducted on computational models of functions relating to the hippocampus from the perspective of computational neuroscience (Banino et al., 2018; Kowadlo et al., 2019; Schapiro et al., 2017; Scleidorovich et al., 2020). Several hippocampus-inspired SLAM methods have been proposed (Milford et al., 2004; Tang et al., 2018; Yu et al., 2019; Zou et al., 2020). Milford et al. (2004) implemented a biologically inspired mapping system, RatSLAM, which is related to place cells in the hippocampus of a rodent.

Some of the existing SLAM functions can be partially associated with each functional cell in the HF. Metric representation in the MEC is approximated by the map coordinate system in SLAM. Grid cells represent spatial metrics to determine coordinate axes (Hafting et al., 2005). Models of grid-cell spatial firing have been proposed in computational neuroscience studies (Zilli, 2012). In contrast, the map representation in numerous SLAM models is either a 2D or 3D Cartesian coordinate system. Particularly, the occupancy-grid map of SLAM can be regarded as a combined representation of grid and border cells. Additionally, in the self-localization model designed for engineering (Ishibushi et al., 2015), Gaussian distributions are arranged at equal intervals, resulting in a representation resembling grid cells. The state variables in SLAM that represent posture are defined by position coordinates and orientation. Therefore, posture is associated with the firing of grid and head-direction cells.

In landmark-based SLAM, which is a time-series state–space model, the landmark position is estimated using a Kalman filter (Montemerlo et al., 2002). This corresponds to the MEC function for processing distal landmark signals obtained from POR (Kuruvilla et al., 2020). In general SLAM, it is assumed that the landmarks are static. Expanding this to dynamic objects is expected to enable the estimation of the position of the movement of other individuals, for example, as a dynamic event that is part of the external context in the environment. This type of expanded SLAM model may be a candidate model for a research report on the activity pattern of place cells in other rats (Danjo et al., 2018).

In the brain, self-localization is performed by theta-phase precession (Hafting et al., 2008). Phase precession is explicitly modeled as a spatial cell model in Zou et al. (2020). However, in the SLAM studies, there are a few models that explicitly implement phase precession, even in brain-inspired models. Hence, it is implemented by different processes that are functionally similar in engineering. In this study, we propose a discrete-event queue, as shown in Section 7.

Episodic memory is an important function both in the hippocampus and robotics. In the computational neuroscience field, models of the hippocampus in episodic memory have been proposed (Hasselmo & Wyble, 1997; Mcnaughton & Neuroscience, 1987; Treves & Rolls, 1994). Several methods incorporating episodic memory have been proposed for robotics (Furuta et al., 2018; Tang et al., 2017; Ueda et al., 2018; Zou et al., 2020). Ueda et al. (2018) proposed a brain-inspired method (i.e., a particle filter on episode) for agent decision-making. An episode that replays in CA3 may be modeled with a generative process. There are various ways to express episodic memory; however, robots can retain temporal transitions with observed events and refer to them subsequently.

5.2. Semantic mapping and spatial concept formation in robotics

In mobile robots, it is essential to appropriately generalize and form place categories while dealing with the uncertainty of observations. Hence, a semantic mapping approach, including semantics of places and objects, has been actively developed (Garg et al., 2020; Kostavelis & Gasteratos, 2015). To address these issues, PGMs for spatial concept formations have been constructed (Hagiwara et al., 2018; Isobe et al., 2017; Katsumata et al., 2020; Taniguchi et al., 2017, 2020a). Taniguchi et al. (2017, 2020a) realized a PGM for online spatial concept acquisition with simultaneous localization and mapping (SpCoSLAM), which conducts place categorization and mapping through unsupervised online learning from multimodal observations. Visual information is used as landscape features reminiscent of spatial view cells (Rolls, 2013). HF is also centrally involved in the formation of place categories, semantic memory, and understanding of the meaning of places by integrating signals from each sensory organ (Buzsáki & Moser, 2013). We consider the aforementioned models as candidates for a cognitive module with functions similar to those of the HF.

SpCoSLAM is a model that arguably imitates some functions of the hippocampus and cerebral cortex. From the viewpoint of computational efficiency and estimation accuracy, Taniguchi et al. (2020a) proposed an inference algorithm that sequentially re-estimates some recent events and accordingly updates global parameters in older observations. Assuming that the training data (i.e., the event-based robotic experience) represent episodic memory and spatial concepts represent semantic memory, their algorithms can sequentially extract concepts from short-term episodic memory to form a semantic memory. Isobe et al. (2017) proposed a model for place categorization using self-position and object recognition results. This model shows that the categorization accuracy is higher when weighing is performed while considering only the objects that are close to the robot, rather than using all objects that are evident from the robot's viewpoint for place categorization. This result is also consistent with the neuroscientific findings of proximal landmarks (Kuruvilla et al., 2020). Furthermore, the hierarchical multimodal latent Dirichlet allocation (Ando et al., 2013) provides a categorical representation of hierarchical locations (Hagiwara et al., 2018). The multi-layered k-means was adopted to extract the hierarchical positional features of a space. Arguably, this corresponds to the hierarchical representation of grid cells in an MEC (Hafting et al., 2005) and is considered a valid model based on neuroscientific findings. Although the aforementioned algorithms and models were not originally inspired by biology or neuroscience, such research is highly suggestive.

5.3. Neural-network models

Deep neural-network models, such as world models (Ha & Schmidhuber, 2018), represent various types of information in the latent state space. In ordinary SLAM, the functional shapes of the distributions (i.e., global parameters) for motion models, which represent state transitions and the measurement models that match observations to a map, are designed and set by humans. Therefore, SLAM problem means the estimation of local latent parameters. In contrast, world models do not have explicit models embedded with prior knowledge. Deep neural-network models learn elements corresponding to global and local parameters in PGMs simultaneously. The estimation of motion and measurement models, including map representations on neural networks, is related to predictive coding.

As computational models for HF, grid cells and their similar spatial representations have been reproduced by deep neural-networks, including long short-term memory (LSTM) or multi-layered recurrent neural networks (RNNs) (Banino et al., 2018; Christopher J. Cueva, 2018; Noguchi et al., 2019). A vector representation similar to that of grid cells is acquired as an internal state by LSTMs, which predicts the current posture from the past velocity and angular velocity with dropouts (Banino et al., 2018). In contrast, from an engineering point of view, models for place representation and spatial concepts using deep neural-networks have been proposed. The room space is learned from the visual-motor experience using two sub-networks comprising a deep auto-encoder and an LSTM (Yamada et al., 2017). By integrating the spatial concept formation model introduced in Section 5.2 with generative adversarial networks, Katsumata et al. (2020) transferred global spatial knowledge from multiple environments to a new environment. These approaches have some suggestive elements that can be interpreted as an HF model.

With the advent of Bayesian deep learning and deep PGMs, it has become possible to discuss neural networks within the framework of PGMs. In particular, by implementing it in the framework of deep PGMs, it is possible to naturally incorporate hierarchical RNNs into a generative model, such as variational auto-encoder and generative adversarial networks. For example, the predictive-coding-inspired variational RNN (Ahmadi & Tani, 2019) is not a model for the hippocampus or navigation task, but a hierarchical RNN-based deep PGM. Deep PGMs are achieved by amortized variational inference, a type of variational inference that introduces functions to transform observation data into parameters of the approximate posterior distribution. Therefore, it is possible to formulate arbitrary variable transformations and nonlinear functions as probability distributions. The framework for maximizing the evidence lower bound by variational inference is equivalent to free energy minimization. Therefore, if the HF models are described by PGMs, they can be naturally connected to the free energy principle (Friston, 2019) and world models (Ha & Schmidhuber, 2018).

6. HF-PGM: Hippocampal formation-inspired probabilistic generative models

We constructed a graphical model of HF and its surrounding regions: HF-PGM. We describe the HF-PGM in association with the HF by interpreting and modeling it using a graphical model. Section 6.1 describes the region of interest, top-level function, and input/output signals. Section 6.2 describes the execution for allocation of generative or inference process in connections. Section 6.3 describes the time-series representation, encoder–decoder representation, and representation associated with BIF in the proposed HF-PGM. Section 6.4 discusses the consistency of the model with scientific knowledge.

6.1. Region of interest, top-level function, and input/output

The region of interest targeted in this study to realize the proposed model includes LEC, MEC, CA1, CA3, DG, Sb, and ParaSb. Additionally, adjacent areas of the region of interest (i.e., areas where the signal can be obtained or a signal is sent) are PreSb, POR, PER, and RSC. The connection relationship of each region is based on Fig. 3. Fukawa et al. (2020) mainly focused on the MEC and hippocampal circuits. In addition, LEC is considered in this study. Hypotheses on LEC functions were derived from several papers and used for model construction. This study assumes that the model works when the agent is walking in an awake state. The top-level function for the assumed region of interest integrates allocentric and egocentric information once and outputs each prediction. The following activities describe the processing inside the model used to realize the top-level function:

(i) Self-localization (path integration and observational correction): Self-localization is performed by path integration and observation-based prediction correction (as explained in Appendix B.2 and Section 4.2). The difference from conventional SLAM is that the prediction is corrected using integrated information from the LEC-variables, and the prediction of the future times is output to other regions. For details, refer to Section 6.3.1.

(ii) Place categorization by integrating allocentric and egocentric information: By integrating the allocentric visual information processed by the MEC and the egocentric visual information processed by the LEC, a place category representing semantic memory about a place is formed (Buzsáki & Moser, 2013). As a PGM, it can be modeled as a multimodal categorization. For details, refer to "integration of information" in Section 6.3.3.

The inputs to the model include variables in POR, PER, and RSC. The internal representations at these parts are treated as observed variables in PGM. The outputs of the model are the predicted values in POR and PER at the next time-step. In PGM, the latent variables and parameters of the conditional distribution are obtained as the estimated values. The definition of each corresponding variable is described in Table 1.

6.2. Executing generation-inference process allocation

We constructed HF-PGM to be consistent with SLAM's PGM based on the generation-inference process allocation procedure (see Section 3.3). The connection between POR and MEC II superficial on the BIF (Fig. 3) can be regarded as a feedforward pathway. An inference process can be allocated to this connection in the PGM. In contrast, the connection between MEC V and POR on the BIF can be regarded as a feedback pathway. A generative process can be allocated to this connection in the PGM. The connections between LEC and PER are assigned in the same way.

There exist limitations to performing generation-inference process allocation for the inside of the hippocampus while considering only the connectivity with the neocortex. Therefore, the engineering formulation of the SLAM modeled as PGM is used as a reference. Partially following the generative process in the PGM of SLAM (see Appendix B.2), the links related to the state transition, motion models $p(x_t \mid x_{t-1}, u_t)$, were assigned as the generative processes.

6.3. Models and operating principle

In this section, we describe the model structure and function of the proposed HF-PGM. We associate variables with each region of the HF. Some anatomical connections are omitted in the PGM for engineering feasibility. Descriptions of global parameters are omitted from the graphical model. In this study, a specific shape is not particularly limited in probability distributions. Only random

Fig. 5. Graphical model representation of HF-PGM as the final form. It is drawn on the BIF, representing the circuit diagram of the HF shown in Fig. 3. The area surrounded by a gray frame is the region of interest targeted in this study. Gray nodes represent observed variables, whereas white nodes represent unobserved latent variables. Black arrows represent the generative process, and dotted arrows represent the inference process. The flat arrow with Δt indicates the generation of the variable in the next time step. Nodes surrounded by gray circles are assumed to be functionally similar and may be treated as the same variable.

variables, their functional definitions, and dependencies between the random variables are assumed. Here, the node does not necessarily have to be a random variable. Signals from outside the HF-PGM module are treated as observed variables, and it is assumed that the observation of each input is converted from raw data into individual high-level features through each processing module; additionally, it enters the HF-PGM. Models may be considered from raw data that can be observed by robots; however, because we intend focus on the top-level function, for the time being, we consider them as extracted features.

The final form of HF-PGM is shown in Fig. 5. HF-PGM was expressed using inference and generative processes based on the generation-inference process allocation. We describe the variables of HF-PGM in Table 1. Next, two types of descriptions in the table are described. Physiological findings include a description of a physiological phenomenon observed at a specific site on the BIF or the function inferred from it. Because there are many uncertain elements in physiological findings related to LECs, the description is omitted here. The function of the components is a description of the computational functions assumed for each component included in the HCD. As a detail of the aforementioned contents with references, we released a pre-screening version of the BRA data[6] as supplementary material (`hf_bra.xlsx`), which describes the navigation functions of the HF.

Sections 6.3.1–6.3.3 explain the flow of changes from the general SLAM models in a step by step manner, as shown in Fig. 5. This implies that the functions of SLAM's PGM can be decomposed and associated based on the anatomical structure of HF. Meanwhile, the functions of the HF-PGM can include the functions of the conventional SLAM (discussed in Section 5.1) and the spatial concept formation models (introduced in Section 5.2).

Fig. 6. Time-series version of graphical model representation of only the MEC side in HF-PGM. The generation and inference processes are drawn simultaneously. The subscript representing time reflects the time of the outside world, in which the observation is obtained. Notably, it does not represent the time of internal processing. Observations up to the current time, t, are obtained.

The operating principle, as an internal process for realizing top-level function in the HF-PGM, is shown below. These also serve as sub-functions subordinate to the top-level function.

6.3.1. PGM representation according to hippocampus and MEC structures

The graphical model is shown in Fig. 6, which is a variant of the PGM of SLAM (see Fig. B.2) and is mapped with reference to the hippocampal and MEC loop structures. This model is an extension of a partially observable Markov decision process. In this model representation, we omit the connections on the LEC side. However, the signal obtained from the LEC side was originally integrated via H_t, which corresponds to the integrated higher-level internal representation. Corresponding to this graphical model of the neural circuits in the hippocampus and MEC, we obtain

Table 1

Description of variables in the HF-PGM. Gray backgrounds are associated with PGM for SLAM, as shown in Fig. B.2. The corresponding variables are listed together in symbol.

Region	Physiological findings	Symbol	Function of components on HCD
CA1 distal	Non-spatial semantic memory	C	Place category (internal representation of visual spatial information)
CA1 proximal	Place cells	r	Position distribution (cluster information regarding positions)
Sb proximal	State at CA1 distal with time delay	C'	Place category at the previous time
Sb distal	State at CA1 proximal with time delay	r'	Position distribution at the previous time
CA3	Pattern completion, information integration	H^{III}	Integrated semantic memory and episodic memory of information from X and Z
DG	Pattern separation	H^{DG}	Integrated semantic memory
MEC II	Grid cells, Border cells, path integration	X^{II} ($\{x_t\}$)	Self-position information, predictive distribution
MEC III	Grid cells, head-direction cells, border cells	X^{III} ($\{x_t\}$)	Self-posture information (position and orientation), observation likelihood
MEC Va,Vb,VI	Prospective speed calculation, feedback to POR	g	Predictive representation of X (Prediction at future time regarding movement/speed amount or posture)
LEC II	–	Z^{II}	Abstraction of information from y^{PER} (transmission of prediction, generation of prediction signal)
LEC III	–	Z^{III}	Abstraction of information from y^{PER} (Observation transmission)
LEC V,VI	Feedback to PER	l	Predictive representation of Z (Prediction at future time from the difference between C' and C)
ParaSb	–	X' ($\{x_{t-1}\}$)	Self-posture information
POR	–	y^{POR} ($\{y_t\}$)	Allocentric visual information (distal distance/landmarks, absolute object positions)
PER	–	y^{PER}	Egocentric visual information (proximal distance/landmarks, relative object positions, object category, landscape information)
RSC	Head direction signal	u^w ($\{u_t\}$)	Rotational speed movement

the following: CA1/Sb (r_{t-1}) → ParaSb (X'_t) & MEC Vb, VI (g_t) + POR/RSC (y^{POR}_t, u_t) → MEC II (X_t) → DG/CA3 (H_t) & CA1/Sb (r_t) → MEC Vb, VI (g_{t+1}) → POR ($\hat{y}^{\mathrm{POR}}_{t+1}$). Some are defined as the same variable, assuming that they have similar functions; however, the regions are different.

HF-PGM differs from the traditional PGM of SLAM when generating predictions of observation for the next cycle. SLAM directly estimates the current self-posture, x_t, from the previous self-posture, x_{t-1}. Furthermore, the generative process of SLAM is the arrow from x_t to y_t at time t. In contrast, HF-PGM generates the self-posture, X_{t+1}, from X_t via variables, such as H_t, r_t, X'_{t+1}, and g_{t+1}.

The higher-level representation of place is denoted as r_t. g_{t+1} is responsible for the prediction of the next time. It is assumed that r_t comprises differential information about the time delay between CA1 and Sb. Hence, the prediction, g_{t+1}, is generated from r_t, and the time-differential information is used. This prediction by the difference calculation is not performed in conventional SLAM.

Additionally, r_{t-1} at the previous time, $t-1$, generates the self-posture, X'_t. X'_t is responsible for conveying the position information of the previous time. The self-position, X_t, is predicted from X'_t and the movement amount, u_t. Subsequently, the self-position, X_t, is corrected by the predicted value, g_t, and the observation, y^{POR}_t.

It is assumed that the predicted value, g_t, is determined by reducing the error so that X_t can be generated consistently with other variables. It is also possible to minimize the prediction error of $\hat{y}^{\mathrm{POR}}_t$, which is generated from the prediction, g_t, and the actual observation, y^{POR}_t: predictive coding (Rao & Ballard, 1999).
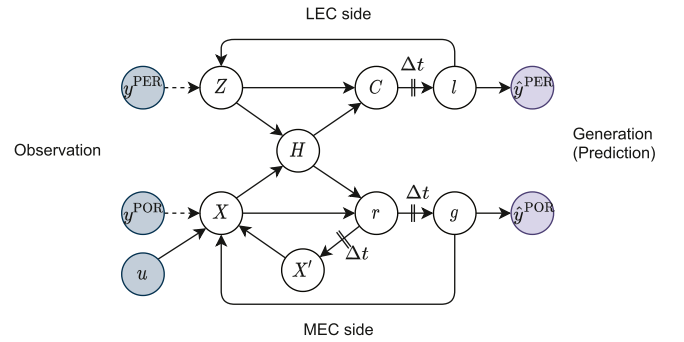
Fig. 7. Encoder–decoder version of graphical model representation in HF-PGM.

Related to the theta-phase precession phenomena in the HF circuit, signals that circulate in the looping circuit in HF require units of discrete-event queues in a current state, which includes both near-past and near-future predictions (see also Section 7). Additionally, given the nature of theta-phase precession, variables that form a large loop may not have self-transition (Butler et al., 2018). Therefore, it is assumed that H_t, r_t, and X_t do not have a direct time self-transition despite having a direct neural connection.

6.3.2. Integrated with LEC

Fig. 7 is a time-omitted representation of the graphical model in Fig. 6 with additional connections on the LEC side. The projection from LEC V, VI l to LEC II, III Z assumes a connection structure similar to MEC. This graphical model representation clearly shows

the result of generation-inference process allocation. Because the HF-PGM in Fig. 7 is compressed in time, a circulation occurs in the generative process. Therefore, the notation for the next time generation process introduced in Fig. B.2 is used. Notably, there exists an arbitrariness of position with which the time progress can be allocated in the PGM loop.

This model has an encoder–decoder structure and can be seen as an extended form of variational encoders–decoders (Bahuleyan et al., 2018) with two modalities and a condition. Because the model structure is similar to encoder–decoder models, the input and output have the same variables with the next time generation process for the inference and generation sides. The signals (y^{PER} and y^{POR}) from PER and POR-parts are processed into Z and X by LEC and MEC, respectively, integrated into H by DC/CA3-part and divided into C and r. Subsequently, $\hat{y}^{\mathrm{PER}}$ and $\hat{y}^{\mathrm{POR}}$ are returned from the predictive information, l and g. During training, the loss function can be designed to match the input to the generated output. The structures of Z to C and X to r are similar to those of the skip connection (i.e., the contracting path) and the U-net (Ronneberger et al., 2015). Furthermore, a loop structure has a recurrent time delay on the MEC side, which is consistent with the anatomy of the MEC and hippocampus. It is suggested that this loop structure is crucial to self-localization (Fukawa et al., 2020).

6.3.3. HF-PGM associated with BIF

Finally, corresponding to the BIF of the HF, Fig. 5 is obtained. The following is a detailed description of each part of the HF-PGM based on Fig. 5.

Input and output: The amount of rotational movement, u_t^w, is transmitted from RSC to MEC III (X^{III}) via PreSb. The translational speed is assumed to be calculated inside the MEC, where the speed cells exist (Hinman et al., 2016; Kropff et al., 2015), because the corresponding region cannot be clearly identified. Hence, the amount of movement, u_t, is obtained by integrating the difference information sent from r and r' with u_t^w. In engineering, it is possible to calculate the speed from the difference information. For example, there are methods such as optical flow. A discussion related to the aforementioned sentences is provided in Appendix C.5. Additionally, this time-difference information, as mentioned in Section 6.3.1, is useful for predicting the internal state and input signal at the next time.[7] POR y^{POR} mainly deals with distal landmarks and distance signals (Kuruvilla et al., 2020). Distant information is useful for self-localization because it can be obtained more robustly than proximal information while moving. PER y^{PER} mainly deals with proximal landmarks and distance signals. This signal is useful to avoid proximal obstacles. Additionally, proximal objects may be useful in forming a place category for the current location (Isobe et al., 2017). Landscape information can also be used to roughly identify the current location (Rolls, 2013). These are similar to the treatments of observations in spatial concept formation, as described in Section 5.2.

Role of LEC/MEC: Generally, in the neocortex, layer III is thought to be responsible for the transmission of observations, and layer II for predictions (Yamakawa, 2020). Given that the LEC/MEC is also part of the neocortex, its role is likely to be preserved. LEC/MEC III receives observation signals from the POR and PER and projects them mainly to CA1 and other areas. LEC/MEC II generates predictive signals and projects them to DG and CA3. Hence, it is considered that LEC/MEC II mainly calculates the

predictive distribution, and LEC/MEC III calculates the observation likelihood. Because the MEC has a coordinate system with grid cells, X represents the robot's posture in the environment. MEC II, III receives the observation, y^{POR}, from the POR and the rotational movement, u_t^w, from the RSC. Egocentric visual information y^{PER} (e.g., proximal object signal Kuruvilla et al., 2020) is sent to Z^{II}, Z^{III} in LEC. The variable l of LEC V, VI is expected to be latent variables that serve as intermediates to send a generative signal to the PER.

Integration of information: Allocentric and egocentric information is integrated into the hippocampus. Hence, DG H^{DG} and CA3 H^{III} form an abstract internal representation of a place that integrates information from the visual information, Z, in LEC and the positional information, X, in MEC. We believe that this internal representation corresponds to the spatial concepts. Furthermore, DG and CA3 are said to have functions of pattern separation and completion, respectively (Bakker et al., 2008). These functions are crucial in the integration of multimodal information. From the PGM perspective, pattern separation may be modeled by parameters that determine the Bayesian prior distribution (Sanders et al., 2020). Specifically, the concentration parameter in Dirichlet process clustering is involved in the automatic determination of the number of clusters (Neal, 2000). In short, allocentric and egocentric information are expected to form a cluster within a unified latent space. Pattern completion can be viewed as the process of regenerating information from DG as defect completion by resampling from a probability distribution. Additionally, error correction of self-position is believed to occur in the hippocampus (Fukawa et al., 2020). By this process, simultaneously integrated information H^{DG} and H^{III} can be used. We assume that place-category formation or information processing occurs in the CA1 distal region. Further, we assume that location-dependent category formation corresponding to place cells occurs in the CA1 proximal region.

6.4. Consistency of model with scientific knowledge

This section discusses the consistency of the model using scientific knowledge and how the outcome can be tested. We believe that the HF-PGM is highly feasible because it is consistent with the anatomical findings of HF, although there may be more detailed variables and dependencies. The main reason is that the original PGM of SLAM already exists. We also follow the path integration of MEC and hippocampus, as discussed in Fukawa et al. (2020). Furthermore, the agreement with BIF and engineering operating principles is described in Section 6.3. Consequently, we successfully map the HF as a PGM by introducing the generation-inference process allocation. In summary, HF-PGM is highly consistent with the brain structure of HF.

The effectiveness of HF-PGM can be verified by solving tasks that can be achieved by integrating LEC (egocentric signal) and MEC (allocentric signal). For example, in situations where localization is difficult in SLAM, we can investigate whether the place category information can distinguish the position. In addition, HF-PGM can be realized as a concrete model by integrating the neural network-based world model and SLAM. Then, further experiments can be conducted to determine whether the latent space representation of the world model, which tends to be unstable in learning, can be complemented by the geometric information of SLAM. We have also provided a pseudo-source code in the form of Neuro-SERKET architecture (Taniguchi et al., 2020) for HF-PGM as supplementary material (`HF-PGM_serket.py`).

The implementation of HF-PGM on a robotic platform is a hot research topic for future studies. Hence, we plan to explore the detailed structure of each model element in subsequent studies. The HF-PGM proposed in this study is a specimen, and the actual

---

[7] As the variables in CA3 and Sb are separated in Fig. 5, it is considered that C and r hold time-difference information as an internal representation. Because r' and C' are said to represent the state having a time delay of r and C in BIF, further studies are required to assess the assumptions on the separability and generation of these variables.
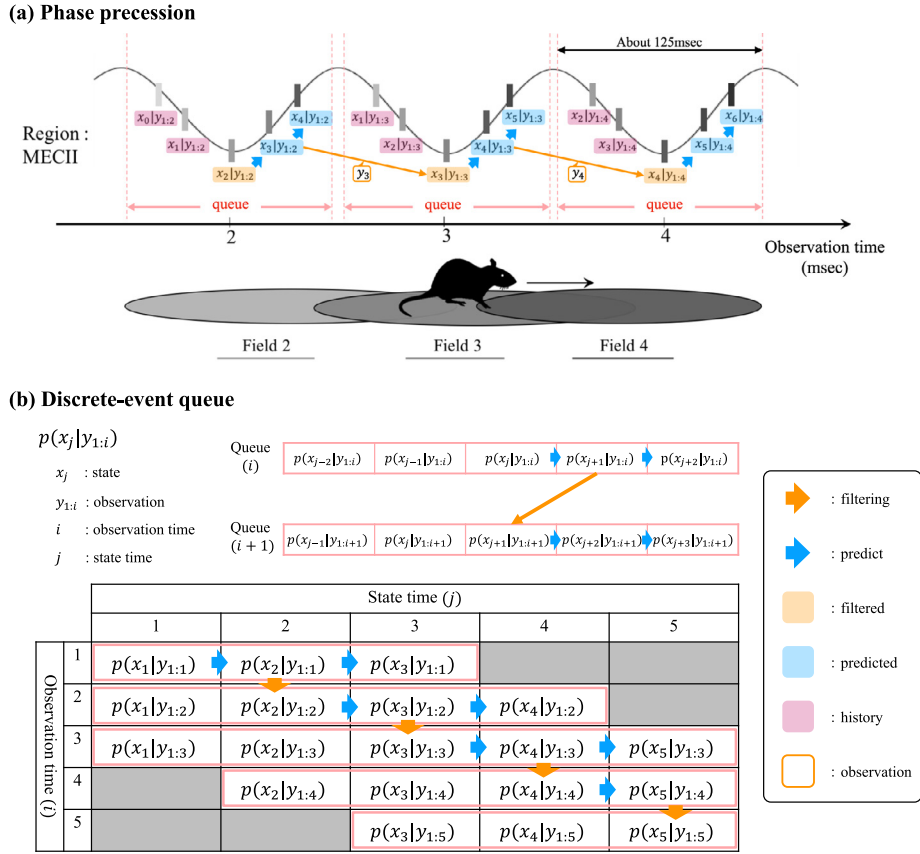
Fig. 8. (a) Phase precession queue assumptions and (b) discrete-event queue processing. (a) shows the information representation in the theta-phase precession of MEC II when the rat moves from left to right. $x_3|y_{1:3}$ represents an estimate of the state, given the observations up to time $t = 3$. Fields 2–4 at the bottom of the rat represent the spatial regions where the three grid cells fire in place fields. The horizontal axis of the bottom table in (b) represents the time of the estimated state, and the vertical axis represents the time of observation. The red box represents the queue. The queue has the estimated/predicted states five times from $t-2$ to $t+2$ at the current time, t. The vertical axis of the table in (b), which is the observation time, corresponds to the horizontal axis in (a), which is the actual time.

implementation may require more concrete discovery in terms of engineering. Therefore, the following issues must be mitigated, including (i) selection of a type for each probability distribution during the generative process, (ii) when performing amortized variational inference, selection of the function shape of the inference model and that of the architecture of the neural networks, and (iii) ensuring a real-time algorithm that includes the learning of global parameters. The aforementioned issues can be solved by model selection and architecture search/optimization in a framework similar to neural architecture search.

7. Abstraction as discrete-event queue

We provide the interpretation of the *phase precession queue assumption*, which is one of the functionalities of the HF as an estimation of the probability distribution in PGM. As explained in Section 6.3.1, a discrete-event queue based on this assumption was constructed to explain the parts that cannot be expressed by the graphical model structure of the PGM alone. The phase precession queue assumption plays an important role in the development of an HCD from BIF. Thus, its relationship with the PGM framework is indirect.

In the HF, the theta-phase precession is known to process the experience by discretizing it and compressing it within a time step, as shown in Fig. 8(a). Herein, stimuli from the external world are sampled at the period of theta waves (8–13 Hz), and it is believed that the present, past, and future events are encoded in phase (Terada et al., 2017). By abstracting the information within one phase as a queue, as shown in this subsection, the calculation

of the queue can be interpreted as filtering the current state, smoothing the past, and predicting the future.

7.1. Introducing discrete-event queue

To comprehensibly model the compressed time process handled by the theta-phase precession, the time granularity of the entire model must be detailed. To avoid too much complexity, we introduce the phase precession queue assumption, which states that the signal held by the theta-phase precession can be regarded as a time queue containing the past, present, and future events at the current time. The assumption is as follows:

Phase precession queue assumption: The observed state compressed within one cycle of theta waves circulating on a pentasynaptic loop circuit can be regarded as a discrete-event queue representing the state sampled at discrete time intervals. Here, the pentasynaptic loop circuit is formed in the hippocampus and MEC by the projection sequence MEC II-DG-CA3-CA1-Sb-ParaSb-MEC II.

There are several reasons why the aforementioned assumption would be considered reasonable. As a neuroscientific finding, HF is expected to be modeled as a discrete-event queue with a finite buffer capacity according to the analysis of large-scale network communication in macaque monkeys (Mišić et al., 2014). Further, in intelligent systems that deal with a world with hidden Markov properties, it is useful to have a discrete-event queue capability to retain the entire observed signal for a short time.

From an engineering perspective, a discrete-event queue can be easily realized as a memory array for the number of time steps to be stored. Here, the memory elements on the same array are considered to have the same meaning, but at different times. However, the neural circuitry of the brain has a restriction that it cannot have more than one representation with the same meaning. Therefore, it seems that we have no choice but to use a method like the theta-phase precession, which compresses information in the time dimension on the same neuron. The fact that PraSb and MEC II on the pentasynaptic loop circuit receive direct projections from the medial septum, which generates theta rhythms, is consistent with the assumption that this circuit is involved in the theta-phase precession.

7.2. Processing for discrete-event queue

From the aforementioned discussion, the information held in the phase precession can be interpreted as a discrete-event queue. If this is the case, the process shown in Fig. 8(b) is performed. Theoretically, the discrete-event queue can be regarded as a sequential estimation problem for the joint posterior distribution in multiple states. The variables are the same as those of the PGM of SLAM, as shown in Fig. B.2.

Table 1 lists the correspondence of variables with HF-PGM. Notably, X_t in the HF-PGM is a variable with an internal representation equivalent to $x_{t-2:t+2}$. Fig. 8(b) shows a table presenting a simplified notation for the discrete-event queue. Each element in the table is a conditional marginal probability distribution in the state, x_j, at time j under the condition of observations up to time i. Here, the control value, u_t, and the integrated information from the LEC are omitted. The probability distribution of the discrete-event queue is shown in Eq. (1).

$$\text{Queue}(t) = p(x_{t-2:t+2} \mid y_{1:t}), \qquad j \in \{i-2 \leq i \leq i+2\}, \quad i = t,$$
$$= \eta \underbrace{p(x_{t+2} \mid x_{t+1})}_{\text{Prediction}} \underbrace{p(y_t \mid x_t)}_{\text{Filtering}} \underbrace{p(x_{t-2:t+1} \mid y_{1:t-1})}_{\int \text{Queue}(t-1)dx_{t-3}}, \quad t \geq 3. \quad (1)$$

Here, we assume that Queue($t − 1$) is calculated at the previous time, $t − 1$. i and j in the formula correspond to those in Fig. 8. $\int \text{Queue}(t-1)dx_{t-3}$ denotes the operation of the integrated-out (i.e., marginalizing) (Murphy, 2012) of x_{t-3} from Queue($t − 1$). As shown in Eq. (1), Queue(t) is a recurrence formula expressed by Queue($t − 1$), and a sequential calculation similar to the Bayes filter is possible.

The discrete-event queue can be interpreted as an algorithm that combines the filtering with the smoother and the prediction (Kitagawa, 2014). This queue calculation is applicable to PGMs of any partially observable Markov decision processes, not just to simple PGMs for SLAM. Refer to Appendix B.1 for the formulae of the predictive/smoothing distributions. In general online self-localization, the belief, $bel(x_t)$, shown in Appendix B.2, which is the distribution when $i = j$, is estimated without using the queue. This is called filtering (Kitagawa, 2014; Thrun et al., 2005). The smoothing distribution can correct past self-position estimates from later observations. SpCoSLAM 2.0 (Taniguchi et al., 2020a) introduced the fixed-lag smoother for sequential and accurate state estimation. The predictive distribution predicts the future self-position from the current state and the learning result of the environment, providing a trajectory that avoids obstacles. Additionally, long-term predictions can be made by repeating predictions ahead by one term. For example, such predictions have already been realized in a map-based motion model (Thrun et al., 2005) and a stochastic model predictive control (Li et al., 2019) for autonomous vehicles. In addition, predictions related to the generation of spatial behavior are discussed in Appendix C.1.

8. Conclusion

We sought to bridge the findings of HFs in neuroscience and SLAM methods in AI and robotics. We summarized the SLAM methods in PGMs and the neuroscientific findings of the HFs and investigated their associations. This paper presents a case study on the framework reported by Yamakawa (2021). The main contribution of this study is the construction of a PGM for the HF that satisfies the evaluation criteria for the BRA design. Specifically, the BIF was designed to be consistent with the anatomy of the HF, and the HCD was then designed by extending the existing SLAM model to be consistent with the structure of BIF. We intend to evaluate the effectiveness of this framework and HF-PGM in future studies. In addition, the generation-inference process allocation solved particular problems regarding the mapping of PGMs to brain circuits. The HF-PGM is significantly different from the previous SLAM models; it integrates LEC and MEC and introduces a discrete-event queue. Such structures were not found in most SLAMs and are very suggestive in engineering modeling.

This study operates as part of the grand challenge of realizing the whole-brain architecture using PGMs (Taniguchi et al., 2022). A whole-brain PGM will be expected to integrate submodules of PGMs for multiple brain regions by using the Neuro-SERKET architecture (Taniguchi et al., 2020). In the case of HF-PGM, visual information obtained from other PGM modules corresponding to its surrounding areas, including the visual cortex, could be connected as observed variables. Other areas of HF connection, such as the prefrontal cortex, were not targeted by HF-PGM in this study. Integration with a PGM module that mimics such areas can be considered in future studies. Other detailed discussions related to open questions and future perspectives are presented in Appendix C.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Construction methodology for brain reference architecture

This section describes the methodology for constructing the BRA (Yamakawa, 2021). Herein, the methodology used to build the BRA model (i.e., the structure-constrained interface decomposition (SCID) method) is introduced.

The whole-brain architecture approach involves iterative and incremental development with the aim of producing a system corresponding to the whole brain capable of general-purpose

problem-solving. This approach is defined as "to create a human-like artificial general intelligence by learning from the architecture of the entire brain".[8] In individual projects, several tasks are solved by assigning them to partial circuits in the brain.

The SCID method is a hypothesis-building method for creating a hypothetical component diagram consistent with the neuroscientific findings. It is required to ensure consistency between neuroscientific findings and engineering feasibility to build brain-inspired models. In current neuroscience research, findings on anatomical structures at the mesoscopic level are obtained on the near-full-brain scale, usually using rodents as the model organism. Therefore, the SCID method can be applied to a wide area of the brain.

To construct the BRA, which includes the design information for brain-inspired AI, anatomical information is first collected, and a BIF is constructed. A BIF is an information flow diagram that describes the mesoscopic-level anatomy of the whole brain. Therefore, it is not intended for a specific task in the environment. BIF, which is a graph, is a basic structure comprising a node (i.e., circuit) and a directed link (i.e., connection). The SCID method mainly designs HCDs that correspond to BIF.

Using the SCID method, an HCD consistent with the anatomical structure in the region of interest is obtained by the following three-step process:

Step 1 While referring to the findings of various studies related to cognitive behaviors of humans and animals, the premise of SCID applicability is established. Specifically, the three processes are performed in parallel.

Step 1-A: The anatomical structure around the region of interest is investigated and registered as a BIF.

Step 1-B: The existence of a component diagram that can realize the input/output is confirmed.

Step 1-C: The valid brain region, the region of interest, and top-level function that it carries are determined.

Step 2 While considering the association between circuits and connections in the region of interest of the BIF, the top-level function is decomposed into detailed functions in as many conceivable patterns as possible. This step enumerates candidate HCDs.

Step 3 HCDs that are logically inconsistent according to scientific findings in various fields (e.g., neuroscience, cognitive psychology, evolution, and development) are rejected. Then, the function of components and the meaning of connections of the remaining HCDs can be assigned to the BIF.

Appendix B. Mathematical preparation of PGMs and SLAM

As preliminary information for computational modeling, we introduce PGMs and SLAM.

B.1. PGMs with graphical model representation

We provide a theoretical introduction and describe the definitions, assumptions, and constraints involved in PGMs. PGMs

represent the process that generates observations as directed acyclic graphs. Graphical models graphically represent the dependencies among random variables in PGMs (see Fig. B.1). Generally, observable variables are represented as gray nodes, whereas unobservable variables (i.e., latent variables) are depicted as white nodes. Global parameters determine the shape of the probability distribution, and local parameters are latent variables that correspond to individual data. Notably, the direction of the arrows does not simply indicate the flow of signal processing; it indicates the generative process of the observation (see Fig. B.1(a)). Basically, the arrows are attached towards the observed data. The vicinity to the root node indicates a high-level latent representation in the brain.

As an example, a graphical model of the variational auto-encoder, which is an auto-encoding variational Bayes (Kingma & Welling, 2014) model, is shown in Fig. B.1.[9] Variational auto-encoder has an inference model,[10] $q_\phi(x|y)$, which is an encoder, and a generative model, $p_\theta(y|x)$, which is a decoder. The flow of signal processing and recognition in the inference model is represented by dotted arrows (see Fig. B.1(b)). When inferring latent variables, an inference model is used to calculate the posterior probability distribution of the latent variables conditioned by the observed values. The inference model in variational auto-encoder is constructed using amortized variational inference (Gershman & Goodman, 2014), which is an approach that introduces functions for efficient approximate inferencing of latent variables. This approach leads to important models that refer to the brain structure (see Section 3.3).

PGMs are separated into the following two phases: (i) the definition of the model structures described in the generative/inference process or graphical model and (ii) the estimating/learning procedure of the posterior/predictive probability distribution and probability. For theoretical details on PGMs, please refer to Murphy (2012).

The state–space models shown in Fig. B.2 and described below are assumed to exhibit a Markov property that adds temporal transitions to latent variable models, including the variational auto-encoder. The state–space models have three types of distributions, depending on the time difference between the state and observation variables. The predictive, filtering, and smoothing distributions are described as follows:

$$\text{predictive distribution}: p(x_t \mid y_{1:t-1}), \tag{B.1}$$

$$\text{filtering distribution}: p(x_t \mid y_{1:t}), \tag{B.2}$$

$$\text{smoothing distribution}: p(x_t \mid y_{1:T}), \tag{B.3}$$

where t is the time of interest, T is the last time, and $(1 \leqq t \leqq T)$. The prediction can be computed from the state transition model. The filtering distribution is typically realized using a Bayesian filter. Smoothing is performed by post-diction, where the past state at time t is updated using the observed signal in the future time, T. It is achieved via data assimilation techniques, such as Kalman smoothing and fixed-lag smoothing, on state–space models (Kitagawa, 2014).

B.2. SLAM

In this section, we describe the theory and classification of SLAM (Thrun et al., 2005) and existing methods based on PGMs. SLAM is a common approach for spatial representation in robotics. A graphical model of SLAM, representing the transition properties

---

[8] As the premise for this definition, the central whole-brain architecture hypothesis was set as "the brain combines modules, each of which can be modeled with a machine-learning algorithm to attain its functionalities, thereby combining machine-learning modules such that the brain enables us to construct a generally intelligent machine with human-level or super-human cognitive capabilities".

[9] This figure is modified from Kingma and Welling (2014).

[10] The inference model is sometimes called the recognition model.
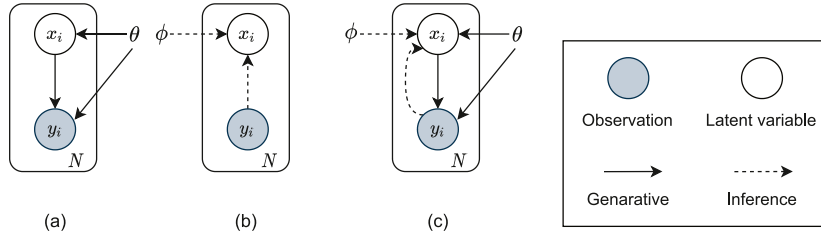
Fig. B.1. Graphical model of latent variable models, such as the variational auto-encoder (Kingma & Welling, 2014). (a) Generative model, (b) inference model, and (c) both generative and inference models. The latent variable denotes x_i, which is a local parameter, and the observation variable denotes y_i. The number of data is N. The index of the data is $i \in \{1, 2, \ldots, N\}$. The global parameter for the generative model is θ, and the global parameter for the inference model is ϕ.
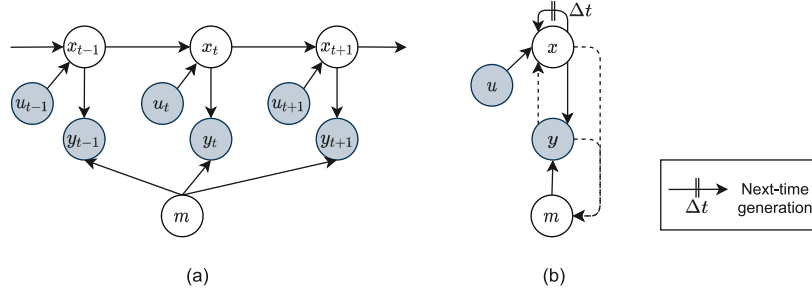
Fig. B.2. Graphical model representations of SLAM. SLAM methods are represented by PGMs based on a partially observable Markov decision process. The self-position denotes x_t, environmental map m, control variable u_t, and observation variable y_t. Global parameters are omitted. (a) Typical drawing of SLAM (Thrun et al., 2005). (b) Compressed drawing of time notation with inference processes, including the original notation of this paper. The flat arrow with Δt indicates the generation of the variable in the next time step.

of the state, control, and observation, is shown in Fig. B.2.[11] This graphical model is commonly referred to as the partially observable Markov decision process.

There are three approaches to solving the SLAM problem: Bayes filtering (Grisetti et al., 2007; Montemerlo et al., 2002), optimization by scan matching (Zhang & Singh, 2017), and pose-graph optimization (Olson et al., 2006). In this study, we primarily focused on the Bayes filter-based online SLAM in PGMs for state–space models. Therefore, the Bayes filtering operation is detailed below. Scan-matching geometrically associates multiple sensor observations with one another. Pose-graph optimization adjusts the positions based on the constraints of a graph representing the trajectory of the robot. Furthermore, a visual SLAM (Uchiyama et al., 2017) involves constructing a 3D map from images.

First, we discuss self-localization, which is a sub-problem of SLAM. The Bayes filter is an algorithm used to estimate the posterior probability distribution of the position with respect to the entire space in the self-localization problem. In probabilistic robotics, it is the principal algorithm for calculating beliefs; however, because it is not a practical algorithm, approximation methods (e.g., Kalman and particle filters) are applied. Belief is a probability distribution calculated on a subjective basis that reflects the robot's internal knowledge of the state of the environment. The belief distribution, $bel(x_t)$, for a state, x_t, can be expressed by Eq. (B.4). This belief distribution represents the posterior probability distribution on the state space conditioned on the observation, $y_{1:t}$, and the control value, $u_{1:t}$, at time t.

$$bel(x_t) = p(x_t \mid y_{1:t}, u_{1:t}). \tag{B.4}$$

The Bayes filter is a sequential process that relies on two important assumptions: prediction and filtering. Eq. (B.5) is the prediction by the motion model, and Eq. (B.6) represents filtering by the measurement update. This prediction is called "dead reckoning" in the navigation field of a voyage and calculates the position by integrating the amount of movement of the robot. The measurement update corrects the error by observation. An iterative update rule is applied to compute $bel(x_t)$ from $bel(x_{t-1})$. The process to estimate belief is shown as follows:

$$\overline{bel}(x_t) = p(x_t \mid y_{1:t-1}, u_{1:t})$$
$$= \int p(x_t \mid x_{t-1}, u_t) bel(x_{t-1}) dx_{t-1}, \tag{B.5}$$

$$bel(x_t) = \eta p(y_t \mid x_t) \overline{bel}(x_t). \tag{B.6}$$

Here, the belief distribution, $bel(x_{t-1})$, is already estimated at a previous time, $t - 1$. The η is the normalization term.

In addition to the aforementioned self-localization, SLAM must simultaneously estimate the environmental map. The PGM-based SLAM methods are estimated based on Bayes filters, such as landmark-based (Montemerlo et al., 2002)/grid-based FastSLAM (Grisetti et al., 2007). FastSLAM estimates joint posterior distribution, as shown below:

$$p(x_{0:t}, m \mid u_{1:t}, y_{1:t})$$
$$= \underbrace{p(m \mid x_{0:t}, y_{1:t})}_{\text{mapping}} \underbrace{p(x_{0:t} \mid u_{1:t}, y_{1:t})}_{\text{self-localization}}, \tag{B.7}$$

where x_t represents the 2-dimensional (2D) coordinates in the Cartesian coordinate system and the head-direction of the agent. y_t represents the distance values to the obstacle obtained from a depth sensor (e.g., laser range finder) or the image features obtained from a camera. The joint distribution of trajectory $x_{0:t}$ and map m can be decomposed by factorization into two processes. The first term represents mapping, and the second represents

---

[11] In this section, because Fig. B.2(b) is shown for reference, the SLAM method that introduces the inference model is not explained. Here, to clarify that circulations with time advancing are acceptable for PGMs, the notation, "*next time generation process*" is introduced. The next time generation process is represented by a double line orthogonal to the generation arrow and the symbol Δt.
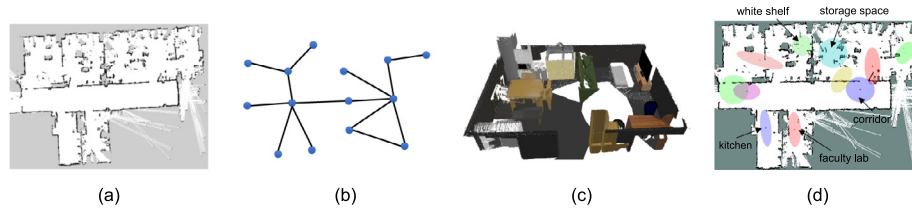
Fig. B.3. Map representation examples in SLAM: (a) Occupancy grid map (Grisetti et al., 2007). (b) Topological map (Blochliger et al., 2018), (c) 3D modeling map (Labbe & Michaud, 2014), and (d) Semantic map (Taniguchi et al., 2020a).

self-localization. Therefore, the algorithm is a sequential iterative SLAM process.

Maps are constructed primarily for use in path-planning and navigation. The map representation based on this method is shown in Fig. B.3: occupancy grid map (Grisetti et al., 2007), landmark map (Montemerlo et al., 2002), 3D modeling of the environment (Labbe & Michaud, 2014; Uchiyama et al., 2017), topological graphs (Blochliger et al., 2018; Mu et al., 2016), and their combination (Choi & Maurer, 2014). The maps can be broadly classified into metric and topological maps. Metric maps consist of feature maps, point-cloud maps, geometric maps, occupancy grid maps, object maps, and semantic maps. Metric maps mainly represent the presence or absence of obstacles or objects in a 2D or 3D Cartesian coordinate system. In topological maps, nodes represent connections in places. The graph structure comprises nodes and edges in a spatial semantic unit, such as a room, or those extracted from the metric map as post-processing. The combination of the metric and topological maps is referred to as a topometric map. A semantic map is a representation that adds semantics to a metric map. More information on semantic maps is presented in Section 5.2.

Appendix C. Open questions and future perspective

Herein, we discuss matters omitted in the HF-PGM alongside unresolved challenges and unclear points.

C.1. Navigation, spatial behavior, and decision-making

Spatial movement and navigation are closely related to HF functions, including spatial cognition and self-localization. The hippocampus is a recognition unit and, therefore, does not directly issue motion commands. Path planning and issuance of motion commands are essential for navigation, and these functions are related to regions connected to the HF (David Poeppel et al., 2020). In engineering, motion generation is related to reinforcement learning and path planning.

Approaches to realizing navigation in robotics are broadly divided into the following two categories.
(i) **Map-based navigation**: One searches for a path to the goal on a map using an objective coordinate system. It is an allocentric coordinate system navigation using SLAM and the grid map. Classical-path planning, such as A* and Dijkstra's algorithms, generally belong to this type of navigation.
(ii) **Visual navigation and vision-based mapless navigation**. The other method (i.e., visual navigation) determines the movement from an image sequence based on an egocentric viewpoint. This navigation is based on the subjective viewpoint signal without a coordinate system.

One could claim that (i) corresponds to the MEC, and (ii) corresponds to the LEC. The aforementioned two types of navigation are the same in predicting and making decisions on future states and actions based on observations. Approach (i) is useful for quickly searching for routes, even at distant goals, because they have an objective perspective and geometric environmental

knowledge similar to those of MEC. In contrast, many navigation types in partially observable Markov decision processes, such as reinforcement learning and vision-and-language navigation, do not use allocentric maps/coordinates, similar to LEC. Approach (ii) enables sequential navigation based on in situ observations; however, it is assumed that it will not be suitable for long-term navigation. Therefore, we project that MEC allocentric information processing is essential for facilitating long-term navigation.

Future prediction is expected to be effective, even when considering the generation of spatial behavior. Humans can decide the subsequent actions in anticipation of future states. For example, one can turn right by predicting a bump into a wall if continuing straight ahead indefinitely. Calculating the predictive value of future position and control is related to the stochastic model predictive control in the (partially observable) Markov decision process models (Li et al., 2019). Furthermore, the connection between active inference (Friston, 2017), control as inference (Levine, 2018), and brain functions, including the HF, is a highly interesting topic for future research.

Long-term prediction can also be formulated as probabilistic inferences by calculating predictive distributions. Taniguchi et al. (2020b) estimated the predictive distribution of future self-position and the amount of motion required from the current position. This method is a map-based navigation based on the control as an inference (Levine, 2018) framework with the same PGM as in SpCoSLAM (Taniguchi et al., 2017). This inference algorithm performs a forward recursive process based on dynamic programming and backward processing from the position having the maximum probability to the current position. It realizes global path planning with speech instruction under spatial concepts acquired from the bottom-up by the robot without setting an explicit goal position.

C.2. Connection with language and meaning

In this study, we constructed a model that mainly represents HF findings in rodents. However, rodents cannot speak; therefore, this model cannot be investigated in terms of its performance in language and meaning processing. Language processing is considered based on human neuroscientific findings. The hippocampal declarative memory system has been identified as a potential key contributor to cognitive functions that require the online integration of multiple signal sources, such as online language processing (Duff & Brown-Schmidt, 2012). In the human brain, the parahippocampal place area, which encodes scenes, is active not only in visual perception but also in the auditory perception of place names (David Poeppel et al., 2020). The parahippocampal place area is thought to be homologous to the POR in rodents. Therefore, this suggests that the input of place names relates to the pathway to LEC, which is responsible for abstraction in place categories.

Considering the connection between the HF and language, it is highly suggestive that robot models can learn the names of

places. These models are also closely related to semantic mapping (Kostavelis & Gasteratos, 2015), which grounds the names of places on a map. Lingodroids[12] realized grounding and sharing of place nouns between mobile robots using the RatSLAM system (Heath et al., 2016; Schulz et al., 2011). In Taguchi et al. (2011) and Taniguchi et al. (2017), robots associated place nouns with specific spatial regions by receiving spoken instruction about the place names from a tutor. Gu et al. (2016) and Sagara et al. (2022) proposed a probabilistic model for grounding the vocabulary of spatial relative positional relationships, including the acquisition of concepts regarding distance and direction. This type of research on the acquisition of spatial language by robots is an important topic in symbol emergence in robotics (Tangiuchi et al., 2019).

Several computational models have been proposed; however, there are very few neuroscientific findings on the association between HF and language. The acquisition of spatial language and its computational modeling are tasks that are to be continued in the future.

C.3. Handling of time and hierarchy

A human can supposedly handle multiple types of hierarchy (Kuipers, 2000). Hierarchies can be divided into three types: (i) temporal, (ii) spatial, and (iii) categorical/conceptual. These hierarchies are not considered in the HF-PGM, but the introduction of a model having a hierarchy is an important direction for future studies. In this section, we introduce possible candidates for the hippocampal computational hierarchy model.

(i) **Temporal hierarchy**. The brain, including the HF, can handle multiple time steps from time compression by theta-phase precession to episodic memory and temporal abstraction time-series processing. Time-series models, such as those in the hidden semi-Markov model (Johnson & Willsky, 2010), LSTM, multiple timescale RNN (Yamashita & Tani, 2010), and predictive-coding-inspired variational RNN (Ahmadi & Tani, 2019) can be considered as models that handle multiple time transitions.

(ii) **Spatial hierarchy** The HF has different spatial resolutions, depending on the area of the brain in the MEC, such as grid cells. Place cells recognize places with a broader and more flexible scope than grid cells. A model with spatial hierarchy may be reproduced using multiple resolution clustering for space, as in Hagiwara et al. (2018) and described in Section 5.2. Topometric maps and hybrid semantic maps can also express spatial resolutions in a hierarchical manner (Pronobis et al., 2017; Rosinol et al., 2020). For example, geometric structures are configured in detail on metric maps, and places are represented abstractly on a topological map.

(iii) **Categories and conceptual hierarchy** Relating to the two aforementioned hierarchies, hierarchy is also present in the representation of place meanings and concepts. Hierarchical clustering using multimodal information is assumed in the computational models. For example, the hierarchical multimodal latent Dirichlet allocation (Ando et al., 2013), pachinko allocation model (Wei & McCallum, 2006), and extended models of variational auto-encoders are applicable. Further, the models introduced in (i) and (ii) may acquire a conceptual hierarchy as internal representations.

C.4. Physicality and sensorimotor system

Anatomical/biological findings in animals other than humans and rodents are also beneficial for developing spatial cognitive systems in robotics. For example, for a drone, it is more advantageous to imitate the brain of a flying animal (e.g., bird or bat). The brain structure has evolved alongside the physicality and sensory organs. When considering the spatial cognitive function of robots and animals, it must be noted that both the internal processes of the cognitive systems and sensorimotor organs (i.e., sensors and actuators) differ.

Actuator: Rodents walk on four legs, humans walk on two legs, birds fly with wings, and dolphins swim with fins. Robots can be wheel-driven with two or four wheels, humanoid robots have two legs, and four-legged robots exist. Drones and underwater vehicles are mainly propeller-driven.

Sensor: Sensors used to facilitate spatial cognition and navigation are not limited to the sensory organs of rodents and humans. Furthermore, some animals use ultrasonic waves and magnetic sensors for navigation. For example, some birds use a magnetic compass to determine their alignments. Robots can be equipped with sensor devices, such as depth sensors, omnidirectional cameras, sonar devices, and global navigation satellite system sensors, that differ from human sensory organs.

Constructing a robot that mimics an animal can serve as a stepping stone to understanding the neural structures required for spatial cognition. Other animals, such as birds and fish, which have similar spatial cognitive functions for constructing cognitive maps, are known to have different HF structures than mammals in evolutionary terms. These topics are discussed in comparative hippocampal science studies (Watanabe & Okaichi, 2008); however, we anticipate that comparisons with robot spatial cognitive models will progress in the future. To this end, we hope that this work will inspire a comparative study of hippocampal functions between mammals and robotics-AIs.

C.5. Mechanism for estimating the absolute speed from proximity visual stimulus

As mentioned in Appendix B.2, from an engineering point of view, the absolute moving speed is measured and input to the SLAM system. In practice, this is realistic because an accurate velocity of the signal can be obtained from mechanisms, such as the rotational speed of the wheels and the integrated value of acceleration by the gyroscope.

However, when humans are on a train, they can sense the train's speed from the scenery passing by; however, they cannot estimate this without windows. Hence, animals detect speed from the visual stimuli of objects that are relatively close to one another, rather than as an absolute speed obtained by engineering methods. Within the scope of our survey, there is no knowledge that the absolute speed of movement is input to HF; the signal on the head-direction is input via the RSC, etc.

Nevertheless, the presence of neural activity associated with *artificial velocity vectors* is known to occur in MEC V with grid cells (Sanders et al., 2015). However, it is not plausible that an artificial velocity vector can be computed within the loop of the MEC and hippocampus because the signal directly input to the MEC via POR is basically information about distant landmarks, which is not suitable for calculating an absolute speed.

However, the proposed model can successfully explain how artificial velocity vectors can be calculated using visual information about proximal objects in the HF. First, the visual information about nearby objects arrives at the DG-CA3 via the PER and the LEC, where it is integrated with allocentric information. Subsequently, the integrated information of the proximal objects is

---

[12] The Lingodroids project includes language learning by mobile robots via spatial language games to construct shared lexicons for places, distances, and directions. https://itee.uq.edu.au/project/lingodroids.

projected to the proximal CA1 and distal Sb, where both are projected to the MEC V, allowing the artificial velocity vector to be computed using a small time-delay difference on the phase precession corresponding to one-time step deference in a discrete-event queue. Measurement of the effect on neural activity related to artificial velocity vector on MEC V by selective inhibition in the PER-LEC pathway can be used to test this mechanism.

Appendix D. List of abbreviations

AI	artificial intelligence
BIF	brain information flow
BRA	brain reference architecture
CA	cornu ammonis area
DG	dentate gyrus
HCD	hypothetical component diagram
HF	hippocampal formation
HF-PGM	hippocampal formation-inspired probabilistic generative model
LEC	lateral entorhinal cortex
LSTM	long short-term memory
MEC	medial entorhinal cortex
ParaSb	parasubiculum
PER	perirhinal cortex
PGM	probabilistic generative model
POR	postrhinal cortex
PreSb	presubiculum
RNN	recurrent neural network
RSC	retrosplenial cortex
Sb	subiculum
SCID	structure-constrained interface decomposition
SERKET	symbol emergence in the robotics tool kit
SLAM	simultaneous localization and mapping
SpCoSLAM	online spatial concept acquisition with simultaneous localization and mapping

Appendix E. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.neunet.2022.04.001.

References

Ahmadi, A., & Tani, J. (2019). A novel predictive-coding-inspired variational RNN model for online prediction and recognition. *Neural Computation*, (June), http://dx.doi.org/10.1162/neco.

Alexander, A. S., Carstensen, L. C., Hinman, J. R., Raudies, F., William Chapman, G., & Hasselmo, M. E. (2020). Egocentric boundary vector tuning of the retrosplenial cortex. *Science Advances*, 6(8), eaaz2322. http://dx.doi.org/10.1126/sciadv.aaz2322, URL http://advances.sciencemag.org/.

Amaral, D. G., & Witter, M. P. (1989). The three-dimensional organization of the hippocampal formation: A review of anatomical data. *Neuroscience*, 31(3), 571–591. http://dx.doi.org/10.1016/0306-4522(89)90424-7.

Ando, Y., Nakamura, T., Araki, T., & Nagai, T. (2013). Formation of hierarchical object concept using hierarchical latent Dirichlet allocation. In *IEEE International conference on intelligent robots and systems* (pp. 2272–2279). http://dx.doi.org/10.1109/IROS.2013.6696674.

Bahuleyan, H., Mou, L., Vechtomova, O., & Poupart, P. (2018). Variational attention for sequence-to-sequence models. In *Proceedings of the international conference on computational linguistics* (pp. 1672–1682). arXiv:1712.08207.

Bakker, A., Kirwan, C. B., Miller, M., & Stark, C. E. (2008). Pattern separation in the human hippocampal CA3 and dentate gyrus. *Science*, 319(5870), 1640–1642. http://dx.doi.org/10.1126/science.1152882, URL http://science.sciencemag.org/.

Banino, A., Barry, C., Uria, B., Blundell, C., Lillicrap, T., Mirowski, P., Pritzel, A., Chadwick, M. J., Degris, T., Modayil, J., Wayne, G., Soyer, H., Viola, F., Zhang, B., Goroshin, R., Rabinowitz, N., Pascanu, R., Beattie, C., Petersen, S., ... Kumaran, D. (2018). Vector-based navigation using grid-like representations

in artificial agents. *Nature*, 557(7705), 429–433. http://dx.doi.org/10.1038/s41586-018-0102-6.

Bermudez-Contreras, E., Clark, B. J., & Wilber, A. (2020). The neuroscience of spatial navigation and the relationship to artificial intelligence. *Frontiers in Computational Neuroscience*, 14(July), 63. http://dx.doi.org/10.3389/fncom.2020.00063, URL https://www.frontiersin.org/article/10.3389/fncom.2020.00063/full.

Bicanski, A., & Burgess, N. (2018). A neural-level model of spatial memory and imagery. *ELife*, 7, 1–45. http://dx.doi.org/10.7554/eLife.33752.

Blochliger, F., Fehr, M., Dymczyk, M., Schneider, T., & Siegwart, R. (2018). Topomap: Topological mapping and navigation based on visual slam maps. In *Proceedings of the IEEE international conference on robotics and automation* (pp. 3818–3825). http://dx.doi.org/10.1109/ICRA.2018.8460641, arXiv:1709.05533.

Butler, J. L., Hay, Y. A., & Paulsen, O. (2018). Comparison of three gamma oscillations in the mouse entorhinal–hippocampal system. *European Journal of Neuroscience*, 48(8), 2795–2806. http://dx.doi.org/10.1111/ejn.13831.

Buzsáki, G., & Moser, E. I. (2013). Memory, navigation and theta rhythm in the hippocampal-entorhinal system. *Nature Neuroscience*, 16(2), 130–138. http://dx.doi.org/10.1038/nn.3304.

Byrne, P., Becker, S., & Burgess, N. (2007). Remembering the past and imagining the future: A neural model of spatial memory and imagery. *Psychological Review*, 114(2), 340–375. http://dx.doi.org/10.1037/0033-295X.114.2.340.

Chen, G., King, J. A., Burgess, N., & O'Keefe, J. (2013). How vision and movement combine in the hippocampal place code. *Proceedings of the National Academy of Sciences of the United States of America*, 110(1), 378–383. http://dx.doi.org/10.1073/pnas.1215834110, URL https://pubmed.ncbi.nlm.nih.gov/23256159/.

Choi, J., & Maurer, M. (2014). Hybrid map-based SLAM with rao-blackwellized particle filters. In *FUSION 2014 - 17th International conference on information fusion*.

Christopher J. Cueva, X.-X. W. (2018). Emergence of grid - like representations by training recurrent neural networks to perform spatial localization. In *Proceedings of the international conference on learning representations* (pp. 1–15).

Danjo, T., Toyoizumi, T., & Fujisawa, S. (2018). Spatial representations of self and other in the hippocampus. *Science*, 359(6372), 213–218. http://dx.doi.org/10.1126/science.aao3898.

David Poeppel, George R. Mangun, & Michael S. Gazzaniga (2020). *the cognitive neurosciences, sixth edition*. The MIT Press.

Deshmukh, S. S., & Knierim, J. J. (2011). Representation of non-spatial and spatial information in the lateral entorhinal cortex. *Frontiers in Behavioral Neuroscience*, 5(OCTOBER), http://dx.doi.org/10.3389/fnbeh.2011.00069.

Doya, K., Ishii, S., Pouget, A., & Rao, R. P. N. (2007). *Bayesian brain: probabilistic approaches to neural coding*. MIT Press, http://dx.doi.org/10.7551/mitpress/9780262042383.001.0001.

Duff, M. C., & Brown-Schmidt, S. (2012). The hippocampus and the flexible use and processing of language. *Frontiers in Human Neuroscience*, (MARCH 2012), http://dx.doi.org/10.3389/fnhum.2012.00069, URL www.frontiersin.org.

Friston, K. (2017). Active inference : A process theory. *Neural Computation*, 49, 1–49. http://dx.doi.org/10.1162/NECO.

Friston, K. (2019). A free energy principle for a particular physics. (pp. 1–148). arXiv:1906.10184.

Fukawa, A., Aizawa, T., Yamakawa, H., & Yairi, I. E. (2020). Identifying core regions for path integration on medial entorhinal cortex of hippocampal formation. *Brain Sciences*, 10(1), http://dx.doi.org/10.3390/brainsci10010028.

Furuta, Y., Okada, K., Kakiuchi, Y., & Inaba, M. (2018). An everyday robotic system that maintains semantic map based on long-term episodic memory. In *Proceedings of the IEEE/RSJ international conference on intelligent robots and systems* (pp. 1–7). IEEE.

Fyhn, M., Molden, S., Witter, M. P., Moser, E. I., & Moser, M. B. (2004). Spatial representation in the entorhinal cortex. *Science*, 305(5688), 1258–1264. http://dx.doi.org/10.1126/science.1099901.

Garg, S., Sünderhauf, N., Dayoub, F., Morrison, D., Cosgun, A., Carneiro, G., Wu, Q., Chin, T.-J., Reid, I., Gould, S., Corke, P., & Milford, M. (2020). Semantics for robotic mapping, perception and interaction: A survey. *Foundations and Trends® in Robotics*, 8(1–2), 1–224. http://dx.doi.org/10.1561/2300000059, arXiv:2101.00443.

Gaussier, P., Banquet, J. P., Cuperlier, N., Quoy, M., Aubin, L., Jacob, P. Y., Sargolini, F., Save, E., Krichmar, J. L., & Poucet, B. (2019). Merging information in the entorhinal cortex: What can we learn from robotics experiments and modeling? *Journal of Fish Biology*, 222(February), http://dx.doi.org/10.1242/jeb.186932.

Gershman, S. J., & Goodman, N. D. (2014). Amortized inference in probabilistic reasoning. In *Proceedings of the annual conference of the cognitive science society, vol. 1* (pp. 517–522).

Gil, M., Ancau, M., Schlesiger, M. I., Neitz, A., Allen, K., De Marco, R. J., & Monyer, H. (2018). Impaired path integration in mice with disrupted grid cell firing. *Nature Neuroscience*, 21(1), 81–93. http://dx.doi.org/10.1038/s41593-017-0039-3.

Giocomo, L. M., Moser, M. B., & Moser, E. I. (2011). Computational models of grid cells. *Neuron*, *71*(4), 589–603. http://dx.doi.org/10.1016/j.neuron.2011.07.023.

Grieves, R. M., & Jeffery, K. J. (2017). The representation of space in the brain. *Behavioural Processes*, *135*, 113–131. http://dx.doi.org/10.1016/j.beproc.2016.12.012.

Grisetti, G., Stachniss, C., & Burgard, W. (2007). Improved techniques for grid mapping with rao-blackwellized particle filters. *IEEE Transactions on Robotics*, *23*, 34–46.

Gu, Z., Taguchi, R., Hattori, K., Hoguro, M., & Umezaki, T. (2016). Learning of relative spatial concepts from ambiguous instructions. In *13th IFAC/IFIP/IFORS/IEA Symposium on analysis, design, and evaluation of human-machine systems* (pp. 150–153). Elsevier.

Ha, D., & Schmidhuber, J. (2018). World models. http://dx.doi.org/10.1016/b978-0-12-295180-0.50030-6, arXiv:1803.10122,

Hafting, T., Fyhn, M., Bonnevie, T., Moser, M. B., & Moser, E. I. (2008). Hippocampus-independent phase precession in entorhinal grid cells. *Nature*, *453*(7199), 1248–1252. http://dx.doi.org/10.1038/nature06957.

Hafting, T., Fyhn, M., Molden, S., Moser, M. B., & Moser, E. I. (2005). Microstructure of a spatial map in the entorhinal cortex. *Nature*, *436*(7052), 801–806. http://dx.doi.org/10.1038/nature03721.

Hagiwara, Y., Inoue, M., Kobayashi, H., & Taniguchi, T. (2018). Hierarchical spatial concept formation based on multimodal information for human support robots. *Frontiers in Neurorobotics*, *12*, 11. http://dx.doi.org/10.3389/fnbot.2018.00011.

Hargreaves, E. L., Rao, G., Lee, I., & Knierim, J. J. (2005). Neuroscience: Major dissociation between medial and lateral entorhinal input to dorsal hippocampus. *Science*, *308*(5729), 1792–1794. http://dx.doi.org/10.1126/science.1110449, URL https://pubmed.ncbi.nlm.nih.gov/15961670/.

Hasselmo, M. E., & Wyble, B. P. (1997). Free recall and recognition in a network model of the hippocampus: Simulating effects of scopolamine on human memory function. *Behavioural Brain Research*, *89*(1–2), 1–34. http://dx.doi.org/10.1016/S0166-4328(97)00048-X, URL https://pubmed.ncbi.nlm.nih.gov/9475612/.

Heath, S., Ball, D., & Wiles, J. (2016). Lingodroids: Cross-situational learning for episodic elements. *IEEE Transactions on Cognitive and Developmental Systems*, *8*(1), 3–14. http://dx.doi.org/10.1109/TAMD.2015.2442619.

Hinman, J. R., Brandon, M. P., Climer, J. R., Chapman, G. W., & Hasselmo, M. E. (2016). Multiple running speed signals in medial entorhinal cortex. *Neuron*, *91*(3), 666–679. http://dx.doi.org/10.1016/j.neuron.2016.06.027.

Ishibushi, S., Taniguchi, A., Takano, T., Hagiwara, Y., & Taniguchi, T. (2015). Statistical localization exploiting convolutional neural network for an autonomous vehicle. In *Proceedings of the 41st IEEE Annual conference of industrial electronics society* (pp. 1369–1375). IEEE, http://dx.doi.org/10.1109/IECON.2015.7392291.

Isobe, S., Taniguchi, A., Hagiwara, Y., & Taniguchi, T. (2017). Learning relationships between objects and places by multimodal spatial concept with bag of objects. In *Proceedings of the international conference on social robotics, vol. 10652 LNAI* (pp. 115–125). Springer, http://dx.doi.org/10.1007/978-3-319-70022-9_12.

Johnson, M. J., & Willsky, A. S. (2010). The hierarchical Dirichlet process hidden semi-Markov model. In *Proceedings of the conference on uncertainty in artificial intelligence* (pp. 252–259). arXiv:1203.3485.

Katsumata, Y., Taniguchi, A., Hafi, L. E., Hagiwara, Y., & Taniguchi, T. (2020). SpCoMapGAN : SPatial concept formation-based semantic mapping with generative adversarial networks. In *Proceedings of the IEEE/RSJ International conference on intelligent robots and systems* (pp. 7927–7934).

Kingma, D. P., & Welling, M. (2014). Stochastic gradient VB and the variational auto-encoder. In *Second international conference on learning representations*.

Kitagawa, G. (2014). Computational aspects of sequential Monte Carlo filter and smoother. *Annals of the Institute of Statistical Mathematics*, *66*(3), 443–471. http://dx.doi.org/10.1007/s10463-014-0446-0.

Knierim, J. J., Neunuebel, J. P., & Deshmukh, S. S. (2014). Functional correlates of the lateral and medial entorhinal cortex: Objects, path integration and local - global reference frames. *Philosophical Transactions of the Royal Society, Series B (Biological Sciences)*, *369*(1635), http://dx.doi.org/10.1098/rstb.2013.0369.

Knill, D. C., & Pouget, A. (2004). The Bayesian brain: the role of uncertainty in neural coding and computation. *Trends in Neurosciences*, *27*(12), 712–719. http://dx.doi.org/10.1016/j.tins.2004.10.007, URL www.sciencedirect.com.

Kostavelis, I., & Gasteratos, A. (2015). Semantic mapping for mobile robotics tasks: A survey. *Robotics and Autonomous Systems*, *66*, 86–103. http://dx.doi.org/10.1016/j.robot.2014.12.006.

Kowadlo, G., Ahmed, A., & Rawlinson, D. (2019). AHA! an -artificial hippocampal algorithm' for episodic machine learning. arXiv, arXiv:1909.10340.

Kropff, E., Carmichael, J. E., Moser, M. B., & Moser, E. I. (2015). Speed cells in the medial entorhinal cortex. *Nature*, *523*(7561), 419–424. http://dx.doi.org/10.1038/nature14622.

Kuipers, B. (2000). The spatial semantic hierarchy. *Artificial Intelligence*, *119*(April 1999), 191–233.

Kuruvilla, M. V., Wilson, D. I. G., & Ainge, J. A. (2020). Lateral entorhinal cortex lesions impair both egocentric and allocentric object–place associations. *Brain and Neuroscience Advances*, *4*, Article 2398212820939446. http://dx.doi.org/10.1177/2398212820939463, URL http://journals.sagepub.com/doi/10.1177/2398212820939463.

Labbe, M., & Michaud, F. (2014). Online global loop closure detection for large-scale multi-session graph-based SLAM. In *Proceedings of the IEEE/RSJ International conference on intelligent robots and systems* (Iros), (pp. 2661–2666). IEEE.

Lever, C., Burton, S., Jeewajee, A., O'Keefe, J., & Burgess, N. (2009). Boundary vector cells in the subiculum of the hippocampal formation. *Journal of Neuroscience*, *29*(31), 9771–9777. http://dx.doi.org/10.1523/JNEUROSCI.1319-09.2009, URL https://www.jneurosci.org/content/jneuro/29/31/9771.full.pdf.

Levine, S. (2018). Reinforcement learning and control as probabilistic inference: Tutorial and review. arXiv preprint arXiv:1805.00909.

Li, N., Girard, A., & Kolmanovsky, I. (2019). Stochastic predictive control for partially observable Markov decision processes with TimeJoint chance constraints and application to autonomous vehicle control. *Journal of Dynamic Systems, Measurement and Control, Transactions of the ASME*, *141*(7), http://dx.doi.org/10.1115/1.4043115, URL http://asmedigitalcollection.asme.org/dynamicsystems/article-pdf/141/7/071007/6416279/ds{_}141_07_071007.pdf.

Madl, T., Chen, K., Montaldi, D., & Trappl, R. (2015). Computational cognitive models of spatial memory in navigation space: A review. *Neural Networks*, *65*, 18–43. http://dx.doi.org/10.1016/j.neunet.2015.01.002.

Madl, T., Franklin, S., Chen, K., & Trappl, R. (2018). A computational cognitive framework of spatial memory in brains and robots. *Cognitive Systems Research*, *47*(September), 147–172. http://dx.doi.org/10.1016/j.cogsys.2017.08.002.

Markov, N. T., Ercsey-Ravasz, M., Van Essen, D. C., Knoblauch, K., Toroczkai, Z., & Kennedy, H. (2013). Cortical high-density counterstream architectures. *Science*, *342*(6158), Article 1238406.

Markov, N. T., Vezoli, J., Chameau, P., Falchier, A., Quilodran, R., Huissoud, C., Lamy, C., Misery, P., Giroud, P., Ullman, S., Barone, P., Dehay, C., Knoblauch, K., & Kennedy, H. (2014). Anatomy of hierarchy: feedforward and feedback pathways in macaque visual cortex. *The Journal of Comparative Neurology*, *522*(1), 225–259.

McNaughton, B. L., Battaglia, F. P., Jensen, O., Moser, E. I., & Moser, M. B. (2006). Path integration and the neural basis of the 'cognitive map'. *Nature Reviews Neuroscience*, *7*(8), 663–678. http://dx.doi.org/10.1038/nrn1932, URL https://pubmed.ncbi.nlm.nih.gov/16858394/.

Mcnaughton, L., & Neuroscience, C. (1987). Hippocampal synaptic enhancement and information storage. *Trends in Neural Sciences*, *10*(10), 408–415.

Milford, M., Wyeth, G., & Prasser, D. (2004). Ratslam: a hippocampal model for simultaneous localization and mapping. In *Proceedings of the IEEE International conference on robotics and automation* (pp. 403–408).

Mišić, B., Goñi, J., Betzel, R. F., Sporns, O., & McIntosh, A. R. (2014). A network convergence zone in the hippocampus. *PLoS Computational Biology*, *10*(12), Article 1003982. http://dx.doi.org/10.1371/journal.pcbi.1003982, URL www.ploscompbiol.org.

Montemerlo, M., Thrun, S., Koller, D., & Wegbreit, B. (2002). FastSLAM: A factored solution to the simultaneous localization and mapping problem. In *Proceedings of the AAAI national conference on artificial intelligence* (pp. 593–598). American Association for Artificial Intelligence.

Mu, B., Giamou, M., Paull, L., Agha-Mohammadi, A. A., Leonard, J., & How, J. (2016). Information-based active SLAM via topological feature graphs. In *2016 IEEE 55th Conference on decision and control* (Cdc), (pp. 5583–5590). http://dx.doi.org/10.1109/CDC.2016.7799127, arXiv:1509.08155.

Murphy, K. P. (2012). *Machine learning: a probabilistic perspective*. Cambridge, MA: MIT Press.

Neal, R. M. (2000). Markov Chain sampling methods for Dirichlet process mixture models. *Journal of Computational and Graphical Statistics*, *9*(2), 249–265, URL http://www.jstor.org/about/terms.html.

Nilssen, E. S., Doan, T. P., Nigro, M. J., Ohara, S., & Witter, M. P. (2019). Neurons and networks in the entorhinal cortex: A reappraisal of the lateral and medial entorhinal subdivisions mediating parallel cortical pathways. *Hippocampus*, *29*(12), 1238–1254. http://dx.doi.org/10.1002/hipo.23145.

Noguchi, W., Iizuka, H., & Yamamoto, M. (2019). Navigation behavior based on self-organized spatial representation in hierarchical recurrent neural network. *Advanced Robotics*, *33*(11), 539–549. http://dx.doi.org/10.1080/01691864.2019.1566088.

O'keefe, J., & Nadel, L. (1978). *the Hippocampus As a Cognitive Map*. Cambridge University Press, http://dx.doi.org/10.5840/philstudies19802725.

Olson, E., Leonard, J., & Teller, S. (2006). Fast iterative alignment of pose graphs with poor initial estimates. In *Proceedings of the IEEE international conference on robotics and automation* (May), (pp. 2262–2269). http://dx.doi.org/10.1109/ROBOT.2006.1642040.

Penny, W. D., Zeidman, P., & Burgess, N. (2013). Forward and backward inference in spatial cognition. *PLoS Computational Biology*, *9*(12), http://dx.doi.org/10.1371/journal.pcbi.1003383.

Pronobis, A., Riccio, F., & Rao, R. P. N. (2017). Deep spatial affordance hierarchy: Spatial knowledge representation for planning in large-scale environments. In *Proceedings of ICAPS 2017 workshop on planning and robotics*.

Rao, R. P., & Ballard, D. H. (1999). Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, 2(1), 79–87. http://dx.doi.org/10.1038/4580, URL https://pubmed.ncbi.nlm.nih.gov/10195184/.

Raudies, F., Brandon, M. P., Chapman, G. W., & Hasselmo, M. E. (2015). Head direction is coded more strongly than movement direction in a population of entorhinal neurons. *Brain Research*, 1621, 355–367. http://dx.doi.org/10.1016/j.brainres.2014.10.053.

Rolls, E. T. (2013). The mechanisms for pattern completion and pattern separation in the hippocampus. *Frontiers in Systems Neuroscience*, 7(OCT), 1–21. http://dx.doi.org/10.3389/fnsys.2013.00074.

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional networks for biomedical image segmentation. In *Medical image computing and computer-assisted intervention* (pp. 234–241). Springer International Publishing.

Rosinol, A., Gupta, A., Abate, M., Shi, J., & Carlone, L. (2020). 3D dynamic scene graphs: Actionable spatial perception with places, objects, and humans. In *Robotics: science and systems*. http://dx.doi.org/10.15607/rss.2020.xvi.079, arXiv:2002.06289.

Sagara, R., Taguchi, R., Taniguchi, A., Taniguchi, T., Hattori, K., Hoguro, M., & Umezaki, T. (2022). Unsupervised lexical acquisition of relative spatial concepts using spoken user utterances. *Advanced Robotics*, 36(1–2), 54–70. http://dx.doi.org/10.1080/01691864.2021.2007168, arXiv:2106.08574.

Sanders, H., Rennó-Costa, C., Idiart, M., & Lisman, J. (2015). Grid cells and place cells: An integrated view of their navigational and memory function. *Trends in Neurosciences*, 38(12), 763–775. http://dx.doi.org/10.1016/j.tins.2015.10.004.

Sanders, H., Wilson, M. A., & Gershman, S. J. (2020). Hippocampal remapping as hidden state inference. *ELife*, 9, 1–31. http://dx.doi.org/10.7554/eLife.51140.

Schapiro, A. C., Turk-Browne, N. B., Botvinick, M. M., & Norman, K. A. (2017). Complementary learning systems within the hippocampus: A neural network modelling approach to reconciling episodic memory with statistical learning. *Philosophical Transactions of the Royal Society, Series B (Biological Sciences)*, 372(1711), http://dx.doi.org/10.1098/rstb.2016.0049, URL https://royalsocietypublishing.org/doi/pdf/10.1098/rstb.2016.0049.

Schulz, R., Wyeth, G., & Wiles, J. (2011). Lingodroids: Socially grounding place names in privately grounded cognitive maps. *Adaptive Behavior*, 19(6), 409–424. http://dx.doi.org/10.1177/1059712311421437.

Scleidorovich, P., Llofriu, M., Fellous, J. M., & Weitzenfeld, A. (2020). A computational model for latent learning based on hippocampal replay. In *Proceedings of the International Joint Conference on Neural Networks*. http://dx.doi.org/10.1109/IJCNN48605.2020.9206824.

Shemyakin, F. N. (1962). Orientation in space. *Psychological Science in the USSR*, 186–255.

Shepherd, G., & Grillner, S. (2013). *Handbook of Brain Microcircuits*. Oxford University Press, http://dx.doi.org/10.1093/med/9780195389883.001.0001.

Taguchi, R., Yamada, Y., Hattori, K., Umezaki, T., Hoguro, M., Iwahashi, N., Funakoshi, K., & Nakano, M. (2011). Learning place-names from spoken utterances and localization results by mobile robot. In *Proceedings of the annual conference of the international speech communication association* (August), (pp. 1325–1328).

Tang, H., Huang, W., Narayanamoorthy, A., & Yan, R. (2017). Cognitive memory and mapping in a brain-like system for robotic navigation. *Neural Networks*, 87, 27–37. http://dx.doi.org/10.1016/j.neunet.2016.08.015.

Tang, H., Yan, R., & Tan, K. C. (2018). Cognitive navigation by neuro-inspired localization, mapping, and episodic memory. *IEEE Transactions on Cognitive and Developmental Systems*, 10(3), 751–761. http://dx.doi.org/10.1109/TCDS.2017.2776965.

Tangiuchi, T., Mochihashi, D., Nagai, T., Uchida, S., Inoue, N., Kobayashi, I., Nakamura, T., Hagiwara, Y., Iwahashi, N., & Inamura, T. (2019). Survey on frontiers of language and robotics. *Advanced Robotics*, 33(15–16), 700–730. http://dx.doi.org/10.1080/01691864.2019.1632223.

Taniguchi, A., Hagiwara, Y., Taniguchi, T., & Inamura, T. (2017). Online spatial concept and lexical acquisition with simultaneous localization and mapping. In *Proceedings of the IEEE/RSJ international conference on intelligent robots and systems* (pp. 811–818). http://dx.doi.org/10.1109/IROS.2017.8202243.

Taniguchi, A., Hagiwara, Y., Taniguchi, T., & Inamura, T. (2020a). Improved and scalable online learning of spatial concepts and language models with mapping. *Autonomous Robots*, 44(6), 927–946. http://dx.doi.org/10.1007/s10514-020-09905-0.

Taniguchi, A., Hagiwara, Y., Taniguchi, T., & Inamura, T. (2020b). Spatial concept-based navigation with human speech instructions via probabilistic inference on Bayesian generative model. *Advanced Robotics*, 34(19), 1213–1228. http://dx.doi.org/10.1080/01691864.2020.1817777, URL https://www.tandfonline.com/doi/full/10.1080/01691864.2020.1817777.

Taniguchi, T., Nakamura, T., Suzuki, M., Kuniyasu, R., Hayashi, K., Taniguchi, A., Horii, T., & Nagai, T. (2020). Neuro-SERKET: Development of integrative cognitive system through the composition of deep probabilistic generative models. *New Generation Computing*, 38(1), 23–48. http://dx.doi.org/10.1007/s00354-019-00084-w, arXiv:1910.08918.

Taniguchi, T., Piater, J., Worgotter, F., Ugur, E., Hoffmann, M., Jamone, L., Nagai, T., Rosman, B., Matsuka, T., Iwahashi, N., & Oztop, E. (2019). Symbol emergence in cognitive developmental systems: A survey. *IEEE Transactions on Cognitive and Developmental Systems*, 11(4), 494–516. http://dx.doi.org/10.1109/TCDS.2018.2867772, arXiv:1801.08829.

Taniguchi, T., Yamakawa, H., Nagai, T., Doya, K., Sakagami, M., Suzuki, M., Nakamura, T., & Taniguchi, A. (2022). A whole brain probabilistic generative model: Toward realizing cognitive architectures for developmental robots. *Neural Networks*, 150(June), 293–312. http://dx.doi.org/10.1016/j.neunet.2022.02.026, arXiv:2103.08183.

Taube, J. S. (2007). The head direction signal: Origins and sensory-motor integration. *Annual Review of Neuroscience*, 30, 181–207. http://dx.doi.org/10.1146/annurev.neuro.29.051605.112854.

Taube, J. S., Muller, R. U., & Ranck, J. B. (1990). Head-direction cells recorded from the postsubiculum in freely moving rats. I. Description and quantitative analysis. *The Journal of Neuroscience*, 10(2), 420–435.

Terada, S., Sakurai, Y., Nakahara, H., & Fujisawa, S. (2017). Temporal and rate coding for discrete event sequences in the hippocampus. *Neuron*, 94(6), 1248–1262.e4. http://dx.doi.org/10.1016/j.neuron.2017.05.024.

Thrun, S., Burgard, W., & Fox, D. (2005). *Probabilistic Robotics*. MIT Press.

Tolman, E. C. (1948). Cognitive maps in rats and men. *Psychological Review*, 55(4), 189–208. http://dx.doi.org/10.1037/h0061626.

Treves, A., & Rolls, E. T. (1994). Computational analysis of the role of the hippocampus in memory. *Hippocampus*, 4(3), 374–391. http://dx.doi.org/10.1002/hipo.450040319.

Uchiyama, H., Ikeda, S., & Taketomi, T. (2017). Visual SLAM algorithms: a survey from 2010 to 2016. *IPSJ Transactions on Computer Vision and Applications*, 9(1), 16.

Ueda, R., Kato, M., Saito, A., & Okazaki, R. (2018). Teach-and-replay of mobile robot with particle filter on episode. In *Proceedings of the IEEE international conference on robotics and automation* (pp. 3475–3481). IEEE, http://dx.doi.org/10.1109/ICRA.2018.8461235.

Wang, C., Chen, X., & Knierim, J. J. (2020). Egocentric and allocentric representations of space in the rodent brain. *Current Opinion in Neurobiology*, 60, 12–20. http://dx.doi.org/10.1016/j.conb.2019.11.005.

Watanabe, S., & Okaichi, H. (2008). *Comparative study of hippocampal functions* (Japanese). (pp. 1–352). Nakanishiya Shuppan, URL http://www.nakanishiya.co.jp/book/b134528.html.

Wei, L., & McCallum, A. (2006). PachInko allocation: DAG-structured mixture models of topic correlations. In *Proceedings of the international conference on machine learning, vol. 148* (pp. 577–584). http://dx.doi.org/10.1145/1143844.1143917.

Witter, M. P., Naber, P. A., Van Haeften, T., Machielsen, W. C., Rombouts, S. A., Barkhof, F., Scheltens, P., & Lopes Da Silva, F. H. (2000). Cortico-hippocampal communication by way of parallel parahippocampal-subicular pathways. *Hippocampus*, 10(4), 398–410. http://dx.doi.org/10.1002/1098-1063(2000)10:4<398::AID-HIPO6>3.0.CO;2-K.

Yamada, T., Ito, S., Arie, H., B, T. O., & Ogata, T. (2017). Learning of labeling room space for mobile robots based on visual motor experience. *1*, In *Proceedings of the international conference on artificial neural networks* (pp. 35–42). Alghero, Sardinia, Italy: Springer.

Yamakawa, H. (2020). Revealing the computational meaning of neocortical interarea signals. *Frontiers in Computational Neuroscience*, 14, 74. http://dx.doi.org/10.3389/fncom.2020.00074, URL https://www.frontiersin.org/article/10.3389/fncom.2020.00074/full.

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. http://dx.doi.org/10.1016/j.neunet.2021.09.004, arXiv:2103.06123.

Yamashita, Y., & Tani, J. (2010). Correction: Emergence of functional hierarchy in a multiple timescale neural network model: A humanoid robot experiment. *PLoS Computational Biology*, 6(10), Article 1000220. http://dx.doi.org/10.1371/annotation/c580e39c-00bc-43a2-9b15-af71350f9d43, URL www.ploscompbiol.org.

Yu, F., Shang, J., Hu, Y., & Milford, M. (2019). NeuroSLAM: a brain-inspired SLAM system for 3D environments. *Biological Cybernetics*, 113(5–6), 515–545. http://dx.doi.org/10.1007/s00422-019-00806-9.

Zhang, J., & Singh, S. (2017). Low-drift and real-time lidar odometry and mapping. *Autonomous Robots*, 41(2), 401–416. http://dx.doi.org/10.1007/s10514-016-9548-2.

Zilli, E. A. (2012). Models of grid cell spatial firing published 2005–2011. *Frontiers in Neural Circuits*, 6(APRIL2012), 1–17. http://dx.doi.org/10.3389/fncir.2012.00016.

Zou, Q., Cong, M., Liu, D., Du, Y., & Lyu, Z. (2020). Robotic episodic cognitive learning inspired by hippocampal spatial cells. *IEEE Robotics and Automation Letters*, 5(4), 5573–5580. http://dx.doi.org/10.1109/lra.2020.3009071.