Accepted Manuscript

International Journal of Neural Systems

Article Title: Enhancement of Hippocampal Spatial Decoding Using a Dynamic Q-Learning Method with a Relative Reward Using Theta Phase Precession

Author(s): Bo-Wei Chen, Shih-Hung Yang, Yu-Chun Lo, Ching-Fu Wang, Han-Lin Wang, Chen-Yang Hsu, Yun-Ting Kuo, Jung-Chen Chen, Sheng-Huang Lin, Han-Chi Pan, Sheng-Wei Lee, Xiao Yu, Boyi Qu, Chao-Hung Kuo, You-Yin Chen, Hsin-Yi Lai

DOI: 10.1142/S0129065720500483

Received: 20 May 2020
Accepted: 21 May 2020

To be cited as: Bo-Wei Chen et al., Enhancement of Hippocampal Spatial Decoding Using a Dynamic Q-Learning Method with a Relative Reward Using Theta Phase Precession, International Journal of Neural Systems, doi: 10.1142/S0129065720500483

Link to final version: [https://doi.org/10.1142/S0129065720500483](https://doi.org/10.1142/S0129065720500483)

This is an unedited version of the accepted manuscript scheduled for publication. It has been uploaded in advance for the benefit of our customers. The manuscript will be copyedited, typeset and proofread before it is released in the final form. As a result, the published copy may differ from the unedited version. Readers should obtain the final version from the above link when it is published. The authors are responsible for the content of this Accepted Article.
ENHANCEMENT OF HIPPOCAMPAL SPATIAL DECODING USING A DYNAMIC Q-LEARNING METHOD WITH A RELATIVE REWARD USING THETA PHASE PRECESSION

Bo-Wei Chen1,†, Shih-Hung Yang2,†, Yu-Chun Lo3, Ching-Fu Wang1, Han-Lin Wang1, Chen-Yang Hsu1, Yun-Ting Kuo1, Jung-Chen Chen1, Sheng-Huang Lin4,5, Han-Chi Pan6, Sheng-Wei Lee2, Xiao Yu7,8, Boyi Qu7,8, Chao-Hung Kuo1,9,10, You-Yin Chen1,3,*, Hsin-Yi Lai7,8,*

1 Department of Biomedical Engineering, National Yang Ming University, No.155, Sec. 2, Linong St., Taipei 11221, Taiwan
2 Department of Mechanical Engineering, National Cheng Kung University, No. 1 University Rd., Tainan 70101, Taiwan
3 The Ph.D. Program for Neural Regenerative Medicine, College of Medical Science and Technology, Taipei Medical University, No. 250 Wuxing St., Taipei 11031, Taiwan.
4 Department of Neurology, Hualien Tzu Chi Hospital, Buddhist Tzu Chi Medical Foundation, No. 707, Sec. 3, Chung Yang Rd., Hualien 97002, Taiwan
5 Department of Neurology, School of Medicine, Tzu Chi University, No. 701, Sec. 3, Zhongyang Rd., Hualien 97004, Taiwan
6 National Laboratory Animal Center, No. 99, Lane 130, Sec. 1, Academia Rd., Taipei 11571, Taiwan
7 Department of Neurology of the Second Affiliated Hospital, Interdisciplinary Institute of Neuroscience and Technology, Key Laboratory of Medical Neurobiology of Zhejiang Province, Zhejiang University School of Medicine, Hangzhou 310029, China
8 College of Biomedical Engineering and Instrument Science, Key Laboratory for Biomedical Engineering of Ministry of Education, Zhejiang University, Hangzhou 310027, China
9 Department of Neurosurgery, Neurological Institute, Taipei Veterans General Hospital, No. 201, Sec. 2, Shipai Rd., Taipei 11217, Taiwan
10 Department of Neurological Surgery, University of Washington, No.1959 NE Pacific St., Seattle, WA 98195-6470, U.S.A.

* laihy@zju.edu.cn
* irradiance@so-net.net.tw

Hippocampal place cells and interneurons in mammals have stable place fields and theta phase precession profiles that encode spatial environmental information. Hippocampal CA1 neurons can represent the animal’s location and prospective information about the goal location. However, the traditional Q-learning (tQ-learning) limits the reward function once the animals arrive at the goal location, leading to unsatisfactory location accuracy and convergence rates. Therefore, we proposed a revised version of the Q-learning algorithm, dynamical Q-learning (dQ-learning), which assigns the reward function adaptively to improve the decoding performance. Firing rate was the input of the neural network of dQ-learning and was used to predict the movement direction. On the other hand, phase precession was the input of the reward function to update the weights of dQ-learning. Trajectory predictions using dQ-learning were compared by the root mean squared error between the actual and predicted rat trajectories. Using dQ-learning, significantly higher prediction accuracy and faster convergence rate were obtained compared with tQ-learning in all cell types. Moreover, combining place cells and interneurons with theta phase precession improved the convergence rate and prediction accuracy. The proposed dQ-learning algorithm is a quick and more accurate method to perform trajectory reconstruction and prediction.

Keywords: place cell, interneuron, dynamical Q-learning, phase precession, adaptive reward function, goal-direction navigation

† Co-first authors
* Corresponding author
1. Introduction

Hippocampal place cells are important for spatial cognition and navigation in mammals.\cite{11,12} Place cells are the basic units of the spatial cognition map; they increase their firing rate whenever the animal enters a firing area or place field.\cite{6,7} Multiple place cells firing within the hippocampus can “map” the environment and provide information on animals’ current location.\cite{8,9,10} In addition, when rodents run through place fields, the firing phase of the corresponding place cells relative to the theta rhythmic oscillations (6–10 Hz) in local field potentials (LFPs) shifts from a later to an earlier phase. This phase shift is called the theta phase precession.\cite{11,12} The place cells encode spatial information by timing spikes to correlate exactly with both the animal’s position in its environment and the phase of the theta rhythm when the animal demonstrates spatial behavioral tasks.\cite{13-15} Therefore, the spatial information obtained from the firing rate and phase precession of place cells can be used to decode or predict the position and construct a movement trajectory.\cite{16,17}

Many studies have demonstrated that decoding spatial information with theta phase precession could reconstruct animal trajectories.\cite{18,19,20} However, using the number of spikes fired on the descending phase of the theta cycles resulted in a less accurate reconstruction of position compared with using the number of spikes fired on the ascending phase of the theta cycles.\cite{15} In fact, Bayesian-based decoding algorithms predict the animal’s position from the ensemble of place cell firing patterns over complete theta cycles.\cite{18,21-25} The prevailing view is that the firing rate of place cells is independent of the animal’s direction of movement.\cite{26,27,28} However, some studies have suggested that place cells exhibit differences in firing rates, which were modulated by the direction of the animal’s head at the same spatial location (head-direction tuning).\cite{29,31} This phenomenon could interfere with the predicted accuracy of position because Bayesian-based decoding algorithms construct the relationships between neural signals and a rat’s position. This problem was solved by choosing place cells with large place fields that cover the entire environment including the goal location, because the firing rates within those place fields were almost the same with various head directions.\cite{32} Unfortunately, those place fields usually uncover the whole environment during goal-directed navigation, which could result in failure to adapt to a complex environment.\cite{34}

Previous studies have used a traditional Q-learning (tQ-learning) algorithm to predict an animal’s direction of travel and avoid the problems associated with head-direction tuning.\cite{35,36} The tQ-learning is a temporal difference-based off-policy reinforcement learning (RL) algorithm that was confirmed using one-step error-driven learning to establish the relationship between firing rate (state) and a rat’s direction (action) with feed-forward neural networks.\cite{37,38} The tQ-learning algorithm collects data about place cell firing rate from the current location to predict a rat’s direction of movement and a reward is subsequently provided; this reward function provides only a single positive reward when the animal arrives at the goal location.\cite{35,36} Moreover, tQ-learning cannot provide any reward when the current location is far from the goal location, thereby failing to adapt to a new environment and causing a slower convergence rate.\cite{39,40}

Thus, the concurrent Q-learning (cQ-learning) algorithm was proposed to solve the problem of reward function.\cite{41} The cQ-learning is defined as a goal-independent method capable of acquiring a reward for every moment, even when the rat is far from the goal location. Although cQ-learning shows higher accuracy in the predicted trajectory, the convergence rate of cQ-learning is slower than that of tQ-learning. Many tQ-learning models have been modified to improve both predictive accuracy and the convergence rate of the reward function, but a significantly faster convergence rate has not been reported to apply the prediction of the animal’s direction. The interneurons in the hippocampus are associated with spatial information and exhibit much larger place fields than place cells. The spatial firings of interneurons are driven by, and show a positive or negative correlation with, place cells.\cite{42-45} Furthermore, positively or negatively correlated interneurons exhibit phase precession.\cite{42} Both interneurons and place cells should contribute to the study of spatial navigation. To our knowledge, the previously mentioned goal-navigation methods, including the Bayesian-based decoding algorithm,\cite{21-23} tQ-learning,\cite{35,36} and cQ-learning\cite{41,46} use only place cells as inputs and do not use information from interneurons. Phase precession has recently been associated with the experience of a rat at destinations. This motivated us to use phase precession as an input for the reward function, enabling a reward at each moment, which might improve the convergence rate and prediction accuracy.

This study was aimed to enhance hippocampal spatial decoding in a rodent model via the proposed dynamical Q-learning (dQ-learning) algorithm during a water-reward task. The neuronal signals of the place cells and interneurons were acquired from the hippocampus to decode the position of a rat when the rat traversed through a maze to obtain a water reward. Furthermore, theta phase precession, which provides a temporal code for trajectory, was adopted to update the dQ-learning algorithm at each time step. Therefore, the navigation model could be immediately adjusted and achieve a faster convergence rate than the tQ-learning, which rewarded the animal only once when it reached the goal location. We evaluated the convergence rates and learning performances of tQ-learning and dQ-learning with different cell types. Indeed, the results demonstrate that dQ-learning improves learning performance and convergence rate and place cells and interneurons with phase precession may provide valuable information to improve the prediction of trajectory.
2. Materials and Methods

2.1. Animal preparation

Three male Wistar rats (Rat 180, Rat 187, and Rat 213; 250–350 g) were housed in an animal center at National Yang Ming University under controlled laboratory conditions (12 h light/dark cycle and room temperature at 22 ± 3°C) and fed twice a day for use in this study. All experimental procedures were carried out in accordance with approved guidelines of the Institutional Animal Care and Use Committee of National Yang Ming University.

2.2. Electrodes implantation and animal surgery

Each microelectrode array was made using eight stainless steel wires (M177390, diameter of 0.002 ft., California Fine Wire Co., Grover Beach, CA, USA) with a horizontal separation of 400 μm. During the implantation, the rats were positioned in a stereotaxic device and anesthetized with a mixture of 40 mg/kg Zoletil 50 (Zolazepam 125 mg/Tiletamine 125 mg, Virbac, Carros, France) and 8 μg/kg Dexdomitor (dexmedetomidine hydrochloride, Pfizer Inc., NY, USA) using intramuscular injections. A scalp incision was made at the midline on the rat’s head to expose the landmark of bregma and lambda suture (Model 900, Kopf Instruments, Tujunga, CA, USA). Then, two microelectrode arrays were respectively implanted into the bilateral hippocampus (3.5 mm posterior and 2.2 mm right- and left-lateral to bregma, 2.2 mm ventral to the cortical surface) and anchored to the skull screws and bone surface with dental cement (type 1 class 1, Hygienic Corp., Akron, OH, USA). Following a recovery period of one week, the rats were trained in a water reward task.

2.3. Apparatus

A schematic diagram of the experimental device is shown in Figure 1A. Neuronal activity was recorded using a Multichannel Acquisition Processor (MAP, Plexon Inc., Dallas, TX, USA). The neural signals were transmitted from the headstage to an amplifier (1,000 ×) at 40-kHz sampling rate and filtered by a bandpass filter (300 Hz to 8 kHz) for spikes and a low-pass filter (250 Hz) for LFPs. Meanwhile, the rat movement trajectories were tracked based on the rat’s head and body, which were colored differently from the

Figure 1. Typical experimental setup of the neural recording system and determination of position coordinates on the routes. (A) The animal movement trajectory and neuronal activity were recorded simultaneously while the animal performed goal-directed spatial navigation after water reward-related lever-pressing training. The behavioral cage (30 × 30 cm, 60 cm in height) with a central square barrier (15 × 15 cm, 37.5 cm in height) was equipped with a top-view digital camera and a lever-controlled pressurized water-supply module. The pressing lever and water plate were set at the diagonal corners of the box at a height of 15 cm. The neural spike and LFPs were acquired via the fixed headstage placed on the head of each rat, which was cabled to a PBX preamplifier and MAP system for further processing. The water reward fixed to the corner with the lever that was diagonally set to the reward site was the primary goal. After pressing the lever, the water-restricted rats reached the reward site by circumnavigating either CW or CCW based on choice. (B) Schematic diagram of the polar coordinate transform for the animal circumnavigating CW and CCW, respectively, in the behavioral cage with an acrylic square bottom. The green dot is the center of the Cartesian coordinates (x_c, y_c). The red and blue dots represent the positions of the rat, which were transformed from Cartesian coordinates (x, y) to the degree of polar coordinates θ.
environment, by the behavioral video tracking system with a frame-rate of 30 frames per second (CinePlex, Plexon Inc., Dallas, TX, USA).

2.4. Animal training for water reward task

After the recovery period, the rats were water-deprived for over 8 h before the training task. Subsequently, the rats were trained in the water reward task, for which they were required to stay in the behavioral apparatus for at least 5 h per day. The behavioral training was performed for 3-5 days without neural recordings. The behavioral apparatus consisted of an acrylic cage (30 cm × 30 cm, 60 cm in height) with a barrier box (15 cm × 15 cm, 37.5 cm in height) in the center of the bottom to create a square-shaped linear track (7.5 cm wide), as shown in Figure 1A. The lever and water plate were 10–15 cm high at the diagonal corners of the behavioral apparatus (left-up and right-down corner) and the pathway was 7 cm wide, limiting the animals to navigation in a clockwise (CW) or counterclockwise (CCW) direction on the square-shaped linear track.

In this study, the water reward task was modified for use in the goal-directed navigation experiment, in which the rats were required to press a lever and traverse in a CW or CCW direction to reach the water plate for the water reward. In addition, the water reward was performed using a 0.25-ml drop per pressure. The trial of the water reward task started when the rats left the lever and ended at the water plate. Because the animals were limited to movement in either the CW or CCW direction, water reward task trials were classified into two directions. The criterion for accomplishment of the water-reward task was met when the rats pressed the lever and reached the water plate. The criterion for successful behavioral training in the water reward task was five consecutive repetitions of lever-pressing and water-drinking. Once this criterion had been achieved, electrophysiological recordings of spikes, LFPs, and movement trajectories were acquired simultaneously, while the animals performed the water reward task. The recording data were collected on the remaining 1.5 h per day, and each rat was examined for a total of 10 days.

2.5. Coordinate transformation

The rat movement trajectory was calculated for each two-dimensional (2D) tracking sample from the projection of the relative position of the head marker onto the horizontal plane. As illustrated in Figure 1B, the rat’s trajectory, specified in Cartesian coordinates, was transformed into polar coordinates. Since the rats were restricted to moving on a narrow track in the behavioral apparatus, the distance from the center of the square-shaped track was simplified to a constant for each degree of the azimuth angle \( \phi \) in the polar coordinate, according to Eq. 1. Only the changes in the azimuth angle \( \phi \) of the polar coordinates were considered while the animals moved in a CW or CCW trajectory on the square-shaped track. According to Eq. 1, the reference angle was presented as a ray from the Cartesian center of the square-shaped track to the right. In this study, the start of the rat’s trajectory was set at the water plate (\( \phi = 135° \)); thus, the new angle \( \bar{\phi} \) was defined as \( \bar{\phi} = 0° \) at the water plate, according to Eq. 2, where the position \((x, y)\) of the animal’s moving trajectory at a certain time \( t \) was represented by the Cartesian components along the \( x \) and \( y \) axes. \( x_c \) and \( y_c \) represented the Cartesian center of the square-shaped track in the behavioral apparatus in Figure 1B. The polar angle \( \phi \) was defined as follows:

\[
\phi = \begin{cases} 
\arctan \left( \frac{y-y_c}{x-x_c} \right), & \text{if } (y-y_c) \geq 0 \text{ and } (x-x_c) \geq 0 \\
\arctan \left( \frac{y-y_c}{x-x_c} \right) + 180°, & \text{if } (y-y_c) > 0 \text{ and } (x-x_c) < 0 \\
\arctan \left( \frac{y-y_c}{x-x_c} \right) + 360°, & \text{if } (y-y_c) \leq 0 \text{ and } (x-x_c) \leq 0 \\
\arctan \left( \frac{y-y_c}{x-x_c} \right) + 225°, & \text{if } (y-y_c) < 0 \text{ and } (x-x_c) > 0 
\end{cases}
\]

\[
\bar{\phi} = \begin{cases} 
\phi - 135°, & \text{if } \phi \geq 135° \\
\phi + 225°, & \text{if } \phi \leq 135° 
\end{cases}
\]

2.6. Classification of place cells and interneurons with their corresponding phase precession

To obtain a spike waveform, spikes were first detected from digitized neural data with the threshold-amplitude detection method and then imported into the unsupervised spike sorting software for discrimination by principal component analysis. To classify different cell types as place cells or interneurons, previous studies used three parameters including the mean firing rate, waveform width of the spike, and the average first moment of the autocorrelogram. The place cells were identified by their lower firing rate (< 8 Hz), as compared with the high firing rate (≥ 8 Hz) of interneurons. The width of the spike waveform was defined as the interval between the positive peaks on the spike waveform—the waveform width for interneurons (0.4 ± 0.02 ms) was shorter than that for place cells (0.45 ± 0.08 ms). Furthermore, the speed of action potential repolarization of place cells was faster than that for interneurons and could be reflected by the average first moment of the autocorrelogram. The average first moment of the autocorrelogram was defined as the mean value of the autocorrelogram for a given cell. A previous study has reported that it was 15 ± 0.22 ms and 25 ± 0.1 ms for place cells and interneurons, respectively. Therefore, these parameters combined could be used to identify place cells and interneurons.

Following the classification of place cells and interneurons, a circular-circular correlation coefficient \( \rho_c \) was used (details in Appendix A). The interneurons with \( \rho_c > 0.7 \) were identified as interneurons with phase precession (abbreviated as interneurons w.\( \theta \)), while those with \( \rho_c < 0.7 \) were identified as interneurons without phase precession (abbreviated as interneurons wo.\( \theta \)). Furthermore,
the direction of rat movement was identified by the direction of the slope (ascending or descending) of the circular regression line (spatial selectivity)\(^{[55]}\) which enabled the recognition of travel in the CW or CCW direction in the behavioral apparatus. The recorded spikes and LFP data were processed with the hippocampal spatial decoder using the dQ-learning algorithm in MATLAB (R2015b, Mathworks Inc., Natick, MA, USA).

### 2.7. Dynamical Q-learning

The dQ-learning algorithm was developed to construct the navigation model and predict the rats’ direction of movement. Different from traditional RL, the algorithm in the present study learned to take actions to imitate the rat’s direction of movement when navigating in an environment so that its trajectory could be accurately reconstructed by accumulating rotational motion. The proposed dQ-learning algorithm learned to choose an action toward a goal (reward) based on the neural signal from the hippocampus (state). dQ-learning was based on a two-layer neural network, as shown in Figure 2.

The neural signal at time \( t \) was the firing rate \( f_j(t) \), which was the input of the two-layer neural network and phase precession. \( p_j(t) \) was the input of the reward function, where \( j = P_1, \ldots, P_n \), \( I_1, \ldots, I_m \), \( P \) and \( I \) represent the place cell and interneuron, respectively, and \( n \) and \( m \) represent the number of place cells and interneurons, respectively. The \( f_j(t) \) and \( p_j(t) \) were calculated every 3 sec\(^{[22]}\) with the time bin of 33 ms, corresponding to the frame-rate of the video tracking system. In Figure 2, the firing rates \( f_j(t) \) were associated to the action units \( a_i \) with weighting \( \psi_{ji} \), where \( i = 1, \ldots, 8 \). The action units, \( a_i \), included eight directions: north (N), north-east (NE), east (E), south-east (SE), south (S), south-west (SW), west (W), and north-west (NW). The weighting, \( \psi_{ji} \), first randomly sampled from a uniform distribution over [0, 1]. Then, the dQ-learning algorithm predicted the rat direction by the maximal Q-value. Note that when the maximal Q-value was zero, the predicted direction was for random movement with a probability of \( P_r = 0.25 \), and for keeping the direction of the previous prediction with the probability of \( 1 - P_r \). When Q values were nonzero, a RL strategy was used and the Q-value was defined as follows:

\[
\Delta Q(t) = \Delta Q(t-1) + \alpha \sum_{i=1}^{n} R_i(t) Q_j(t-i, \alpha) - \gamma Q(t)
\]

where \( \Delta Q(t) \) is the change in the Q-value at time \( t \), \( \Delta Q(t-1) \) is the previous change in the Q-value, \( \alpha \) is the learning rate, \( R_i(t) \) is the reward at time \( t \), \( Q_j(t-i, \alpha) \) is the Q-value at time \( t-i \) for the action selected at time \( t \), and \( \gamma \) is the discount factor.
\[
Q(f(t), a_i) = \frac{\sum_{j=1}^{m_i} \Phi(f_j(t)) \psi_{ij}}{\sum_{j=1}^{m_i} \Phi(f_j(t))}
\]

where \( \Phi(f_j(t)) \) was an indicator function determined as follows:

\[
\Phi(f_j(t)) = \begin{cases} 
1, & f_j(t) > 0.5 \\
0, & \text{otherwise}
\end{cases}
\]

The averaging Q-learning rule\(^{[36]}\) was used to update the weighting \( \psi_{ij} \) of the action that was actually taken, \( a_i \), at the time step \( t-1 \), which is updated as shown below:

\[
\psi_{ij} = \psi_{ij} + \alpha (R_j(t) + \gamma \cdot \max Q(f(t), a_i) - \psi_{ij}) \Phi(f_j(t))
\]

\[
i^* = \arg\max_i Q(f(t-1), a_i)
\]

where \( \alpha = 0.7 \) was the learning rate and \( \gamma = 0.7 \) was the discounting factor.\(^{[38]}\) The weight was updated by the winner-takes-all mechanism\(^{[39]}\). The weight connected to the action unit, \( a_i \), and the action of which was actually taken at the time step \( t-1 \), wins the learning chance. Furthermore, only the weight with a firing rate of the \( j \)th place cell that was \( >0.5 \) was updated, because low firing rates might provide a larger prediction error in position. \( R_j(t) \) was a reward function and it was updated as follows:

\[
R_j(t) = \begin{cases} 
1, & \text{if } \text{sign}(p_j(t)) = \text{sign}(\Delta \theta(t-1)) \\
0, & \text{if } \text{sign}(p_j(t)) \neq \text{sign}(\Delta \theta(t-1)) \\
-1, & \text{failed predicted movement}
\end{cases}
\]

\[
\text{sign}(p_j(t)) = \begin{cases} 
1, & \text{if } p_j(t) > 0 \\
-1, & \text{if } p_j(t) < 0
\end{cases}
\]

where \( p_j(t) \) was phase precession, which was quantified through a phase range and defined as a phase shift during a time window. To compare the phase precession in two directions, phase precession was calculated as the difference between the theta phase at the minimum polar coordinate and that at the maximum polar coordinate of the time window (details in Figure 2A and Appendix A). The \( \Delta \theta(t-1) \) was a predicted rotational motion and was different from actual rat’s rotational motion. The \( \Delta \theta(t-1) \) was indirectly determined by the maximal Q-value as shown in Figure 2 and was calculated by the angular displacement of polar coordinates as \( \Delta \theta(t-1) = \theta(t-1) - \theta(t-2) \). The CW direction corresponding to the sign of rotation motion \( \Delta \theta(t-1) = -1 \) and the CCW direction corresponded to the sign of rotation motion \( \Delta \theta(t-1) = +1 \). Once the update for weighting \( \psi_{ij} \) was complete, the rat’s direction of movement was estimated, based on the direction of the neurons with the maximal Q-value, as shown in Figure 2. The rats’ movement trajectory was updated in terms of its Cartesian coordinates \( x(t), y(t) \), as follows:

\[
x(t) = x(t-1) + \Delta x + \sigma_x \\
y(t) = y(t-1) + \Delta y + \sigma_y
\]

where \( \sigma_x \) and \( \sigma_y \) were noise sampled from a uniform distribution over \([-0.3, 0.3]\), and \( \Delta x \) and \( \Delta y \) constituted a constant step size \( s = 1.5 \) cm. The \( \Delta x \) and \( \Delta y \) were calculated as the model with the lowest BIC because the lower BIC value indicated that the

2.8 Performance evaluation and statistical analysis

To investigate and compare the performance of the different neuronal ensembles, they were divided into five groups according to the various combinations of cell types, as follows: 1) place cells only (P group), 2) interneurons only (IN group), 3) place cells and interneurons with phase precession (P & INW group), 4) place cells and interneurons without phase precession (P & INWo group), and 5) place cells and interneurons (P & IN group). For each group, the experimental trials were divided into a training set (the data from the experimental trials during the initial 7 days), which was used to determine the convergence state of the navigation model, and a testing set (the data from the experimental trials during the remaining 3 days), which evaluated the performance of dQ-learning.

To determine convergence state of the navigation model (detailed in Appendix B), a previous study used the mean Q-value, which was obtained from the training set, to indicate the convergence state.\(^{[38]}\) The training convergence occurred when the average error, which is calculated as \( |Q_{\text{mean maximal}} - Q_{\text{ground truth}}| / Q_{\text{ground truth}} \) \(^{[38]}\), between the mean maximal Q-value \( (Q_{\text{mean maximal}}) \) and ground truth \( (Q_{\text{ground truth}}) \), which is defined as the mean maximal Q-value when it is stable, was less than 8% among three rats. Once the dQ- and \( \sigma \)-learning algorithms were in the faster convergence state\(^{[57]}\), their predictive performances could be compared. An additional decoding method, called the approximate Gaussian mixture model (GMM) (GMM)\(^{[38]}\), was used to compare the decoding performance of \( \sigma \)-learning and dQ-learning. The approximate GMM is a point process filter algorithm that uses a mixture of Gaussian models for posterior filter distribution at each time step. Furthermore, it distintively treats spike and non-spike intervals to decode a rat’s position based on hippocampal place cell and interneuron spiking. The optimization of the approximate GMM was computed by the expectation–maximization (EM) algorithm along with the Bayesian information criterion (BIC)\(^{[56, 60]}\) to estimate model parameters (detailed in Appendix C). The optimal approximate GMM was determined as the model with the lowest BIC because the lower BIC value indicated that the
model properly approximated the neural spike data. To test the performance of a decoder and to evaluate the performance accuracy, the root mean squared error (RMSE) between the rat’s actual and predicted trajectories along the X-axis (RMSEx) and Y-axis (RMSEy) was computed. Because RMSE has been widely used to measure the difference between true and estimated trajectories in the field of the brain-machine interface. Therefore, using RMSE is appropriate to represent the decoding performance.

To statistically analyze the comparison of the predictive ability of dQ-learning with Q-learning and approximate GMM, a non-parametric Kruskal-Wallis test followed by Dunn’s post hoc test was applied. The significance level was *p < 0.05, **p < 0.01, and ***p < 0.001, for the Kruskal-Wallis H test, with multiple-comparison performed by Dunn’s test. Data were analyzed using SPSS version 20.0 (SPSS Inc., Chicago, IL, USA). Results were presented as mean ± standard error of the mean (SEM).

3. Results

To investigate whether the enhancement of hippocampal spatial decoding with the dQ-learning method was effective in goal-direction navigation, experimental data were recorded from three rats (Rat 180, Rat 187, and Rat 213). Table 1 lists the numbers of neurons for each rat and the neurons were classified into place cells (15 cells in total) and interneurons (14 cells in total), according to mean firing rate, average waveform width, and average first moment of the autocorrelogram. Meanwhile, the interneurons were further divided into two subtypes, according to whether they exhibited phase precession or not. Table 2 lists the numbers of trials, including CW and CCW for the training and testing sets.

Table 1. The numbers of neurons for the three subjects.

<table>
<thead>
<tr>
<th>Subject</th>
<th>Number of Place cell</th>
<th>Testing set</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>w.θ</td>
<td>wo.θ</td>
</tr>
<tr>
<td>Rat 180</td>
<td>7</td>
<td>4</td>
</tr>
<tr>
<td>Rat 187</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>Rat 213</td>
<td>4</td>
<td>2</td>
</tr>
</tbody>
</table>

Table 2. The numbers of trials for the three subjects. The experimental trials were divided into training and testing sets, which were further divided into CW and CCW routes.

<table>
<thead>
<tr>
<th>Subject</th>
<th>Number of Trial</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Training set</td>
</tr>
<tr>
<td></td>
<td>CW</td>
</tr>
<tr>
<td></td>
<td>CW</td>
</tr>
<tr>
<td>Rat 180</td>
<td>43</td>
</tr>
<tr>
<td>Rat 187</td>
<td>34</td>
</tr>
<tr>
<td>Rat 213</td>
<td>22</td>
</tr>
</tbody>
</table>

3.1. Identification of place cells and interneurons in the hippocampus

The spike train and LFP in the CA1 pyramidal layer were recorded while the rats performed a water reward task.
Supplemental Video (https://youtu.be/OawFUZUpDvY) shows that two place cells fired when the rat ran through an area. Figure 3A indicates the movement trajectories, spike train, theta band LFP oscillations, spike waveforms, and autocorrelograms of the place cells and interneurons from Rat 180. The place cells from this rat exhibited a lower firing rate (7 Hz) than the interneurons (18 Hz for interneurons w.θ and 24 Hz for interneurons wo.θ). The width of the place cell spike waveform (0.67 ms) was longer than that of the interneurons (0.34 ms for interneurons w.θ and 0.38 ms for interneurons wo.θ). The average first moment of the autocorrelogram for the place cells (10 ms) was shorter than that for the interneurons (21 ms for interneurons w.θ and 26.00 ms for interneurons wo.θ). Figure 3B displays a three-dimensional (3D) scatter plot used to identify the place cells (red dot, 15 cells) and interneurons (blue dot: 8 cells w.θ, and green dot: 6 cells wo.θ) in all three of the rats used in this study. Again, the place cell firing rate (5.25 ± 0.11 Hz) was lower than that of the interneurons (26.69 ± 0.34 Hz) and the width of the place cell spike waveform (0.68 ± 0.01 ms) was longer than that of the interneurons (0.36 ± 0.06 ms). The average first moment of the autocorrelogram of the place cells (11.00 ± 0.12 ms) was shorter than that of the interneurons (26.00 ± 0.17 ms). Therefore, the place cells and interneurons could be distinguished by the firing rate, spike waveform width, and the average first moment of the autocorrelogram.

3.2. Spatial selectivity and theta phase precession

The first-column of Figure 4A and 4B shows the mean firing rate as a function of position in the Cartesian coordinate system (2D place field). The second-column panels show the theta phase precession in a CW direction, where the colors represent the mean firing rate as a function of the positions in the polar coordinate system and theta phase. The white dashed line indicates the best fit based on circular-linear regression. The fitting lines of the place cell and interneuron w.θ show a positive slope. (B) Examples of theta phase dynamics in the CCW direction (organized as in A). The fitting lines of the place cell and interneuron w.θ indicate a negative slope. (C) Fitting lines of all cells from the three rats moving in the CW direction. (D) Fitting lines of all cells from the three rats moving in the CCW direction. (E) Distributions of the circular-circular correlation coefficient for the correlation of theta phase with the position in the polar coordinate system for place cells (red, 15 neurons), interneurons w.θ (blue, 8 neurons), and interneurons wo.θ (green, 6 neurons) in the CW direction. (F) Distributions of the circular-circular correlation coefficient in the CCW direction.
system (i.e., 2D place field) for place cells and interneurons from Rat 180 moving in the CW and CCW directions, respectively. The second column of Figure 4A and 4B shows theta phase precession in the CW and CCW directions, respectively. The theta phase precession was quantified using a circular-linear regression method. The slopes (theta phase / polar coordinate) estimated by the circular-linear regression (white dashed lines) of the place cells and interneurons were 7.50 deg/deg (CW direction) and −4.50 deg/deg (CCW direction) and interneurons w.θ were 1.13 deg/deg (CW direction) and −1.05 deg/deg, respectively. No consistent slope was found for interneurons w.θ. All of the slopes from each neuron in all the trials for different cell types were given for the CW and CCW directions (Figure A1 and Figure A2 in Appendix A). The mean slopes of the place cells were 8.54 ± 2.67 deg/deg (CW direction) and −6.68 ± 2.92 deg/deg (CCW direction) and interneurons were 1.81 ± 0.61 deg/deg (CW direction) and −1.63 ± 0.79 deg/deg (CCW direction). As a result, the place cells and interneurons revealed spatial selectivity with positive slopes for the CW direction and negative slopes for the CCW direction. Moreover, the distribution of the circular-circular correlation coefficient between the theta phase and regression line in the CW (Figure 4C) and CCW (Figure 4D) directions was calculated. Compared with the circular-circular correlation coefficient of interneurons w.θ (0.43 ± 0.16 [CW] and 0.48 ± 0.12 [CCW]), place cells (0.87 ± 0.14 [CW] and 0.83 ± 0.11 [CCW]) showed a higher correlation (detailed results in Figure A2 in Appendix A).
3.3. Training progress of the proposed dynamical and traditional Q-learning methods

To evaluate the training progress of both dQ-learning and tQ-learning, the convergence rate of training progress was calculated by trial between the mean maximal Q-value and ground truth. Our study found the mean maximal Q-value was stable at the 20th trial that was defined as the ground truth. In the comparison of the convergence rate of the learning processes, the mean maximal Q-value and RMSE were tracked during each training trial (Figure 5). The mean maximal Q-value corresponded to the best-predicted direction for each of the given firing rates and the quality of the prediction performance was estimated. Both the RMSE$x$ and RMSE$_y$ values decreased when the mean maximal Q-value increased. The results show that the convergence state for dQ-learning was achieved at the 14th trial in the P group, IN group, and $P & INw.θ$ group, the 15th trial in the $P & INwo.θ$ group, and the 19th trial in the $P & IN$ group. The convergence state for tQ-learning was achieved at the 19th trial in all groups. Moreover, the convergence state of the average errors for dQ- and tQ-learning was attained in groups $P$ ($2.85 \%$ and $2.03 \%$), $IN$ ($3.91 \%$ and $6.05 \%$), $P & INw.θ$ ($4.93 \%$ and $4.06 \%$), $P & INwo.θ$ ($5.14 \%$ and $2.30 \%$), and $P & IN$ ($1.58 \%$ and $4.60 \%$), respectively. The mean maximal Q-value, RMSE$x$, and RMSE$_y$ of the convergence state are provided in Table 3. In summary, the convergence rate for dQ-learning was faster than that for tQ-learning in groups $P$, $IN$, $P & INw.θ$, and $P & INwo.θ$.

3.4. Performance evaluation of the dQ-learning, tQ-learning and approximate GMM methods

A comparison of decoding performance from proposed dQ-learning, tQ-learning and approximate GMM was made based on using the predicted trajectories from Rat 180, Rat 187, and Rat 213 in the testing set for different combinations of cell types. In Figure 6, tQ-learning and approximate GMM methods showed larger deviation between predicted and real trajectories as compared with dQ-learning in groups $P$, $IN$, $P & INw.θ$, $P & INwo.θ$, and $P & IN$. The deviations in the $P & INw.θ$, $P & INwo.θ$, and $P & IN$ groups were lower than those in the $P$ and $IN$ groups, indicating that the performance improved when the number of cell types increased.

RMSE$x$ and RMSE$_y$ were calculated to evaluate the effects of cell types for dQ-learning, tQ-learning and approximate GMM methods, as shown in Figure 7. In all groups, dQ-learning showed a significantly smaller RMSE as compared with the methods of approximate GMM and tQ-learning. The lowest RMSE$x$ for dQ-learning (Rat 180: $2.90 \pm 0.11$ cm, Rat 187: $2.88 \pm 0.03$ cm, and Rat 213: $2.96 \pm 0.15$ cm) and tQ-learning (Rat 180: $4.26 \pm 0.77$ cm, Rat 187: $11.97 \pm 1.30$ cm, and Rat 213: $12.81 \pm 2.30$ cm) were observed in the $P & INw.θ$ group. Furthermore, the lowest RMSE$y$ for dQ-learning (Rat 180: $3.03 \pm 0.08$ cm, Rat 187: $2.96 \pm 0.15$ cm, and Rat 213: $2.88 \pm 0.13$ cm) and tQ-learning (Rat 180: $3.62 \pm 0.19$ cm, Rat 187: $11.19 \pm 1.22$ cm, and Rat 213: $12.93 \pm 1.25$ cm) were found in the $P & INw.θ$ group, indicating that the interneuron with phase precession could improve the decoding performance than the interneuron without phase precession for both dQ-learning and tQ-learning. Moreover, the smallest RMSE was obtained as use of the $P & IN$ group for the approximate GMM. These results indicated that the performance of approximate GMM was not improved by the interneuron with phase precession. Additionally, as compared with approximate GMM method, the tQ-learning showed a significantly smaller RMSE in $P$, $P & INw.θ$, $P & INwo.θ$, and $P & IN$ groups while the dQ-learning showed a significantly higher RMSE in the $IN$ group of Rat 187 and Rat 213. The IN group of Rat 187 and Rat 213 exhibited a minimum number of neurons ($N = 3$). The result revealed that the decoding performance was better using approximate GMM than that using tQ-learning when the number of neurons was small. Nevertheless, the dQ-learning exhibited the smallest RMSE in the IN group of Rat 187 and Rat 213, indicating that dQ-learning showed the best decoding performance for small size of neuron number.

![Figure 6. Predicted trajectories of dQ-learning, tQ-learning, and approximate GMM. Real trajectories (black solid line) and predicted trajectories using dQ-learning (blue dashed line), tQ-learning (red dashed line), and approximate GMM (green dashed line) in five data groups.](image-url)
3.5. Data Availability

All data files are available from public repositories. Specifically, the MATLAB code used to predict the animal’s position and the data necessary for decoders are available at Zenodo repository (https://doi.org/10.5281/zenodo.3724076).

4. Discussion

The main result of this study was that the use of dQ-learning could appropriately predict and reconstruct a rat’s trajectory based on data about the place cells and interneurons. We found that using place cells and interneurons with phase precession resulted in better accuracy with both the dQ- and tQ-learning algorithms and dQ-learning presented a smaller RMSE and faster convergence rate compared with tQ-learning.

4.1. Phase dynamics in CA1 interneurons

Our results showed that using the mean firing rate, waveform width of the spikes, and average first moment of the autocorrelogram realized effective identification and classification of cell types. Some interneurons exhibited a phase precession that was consistent with the previous studies.[55, 62-64] Although the interneurons could be distinguished from the place cells, it was difficult to differentiate between the interneurons with and without phase precession. Therefore, we used circular-linear regression to identify the interneurons with phase precession. The interneurons \( w, \theta \) showed similar theta phase dynamics to place cells, but the slope of interneurons \( w, \theta \) was smaller than that of place cells, indicating that the theta phase shifting of place cells was stronger than that of interneurons. In addition, the results showed that place cells were active over one theta cycle, whereas the interneurons \( w, \theta \) were active over two theta cycles, which has been shown in previous studies.[55, 62-64] Previous studies have demonstrated that the phase precession of interneurons is driven by the place cells.[64-67] Therefore, we suggested that the phase precession of interneurons might be a summarized series of place cells resulting in a phase precession of interneurons that can surpass 360°.

4.2. The faster convergence rate of dQ-learning

In this study, the convergence rate of dQ-learning was faster than that of tQ-learning. Both dQ- and tQ-learning were model-free and used RPE determined by the reward function to update the model.[66] The reward function of tQ-learning was only applied once when the animal reached the goal location. With hybrid reward architecture (HRA), the reward function of dQ-learning was updated at each time step using the phase precession and the predicted rotational motion. Therefore, dQ-learning could decide more meticulously which cell to be rewarded and drive the model directly to enable the dQ-learning algorithm to improve the convergence rate.

Figure 7. Comparison of the performance of dQ-learning, tQ-learning, and approximate GMM. The RMSE in X-axis and Y-axis of the dQ-learning, tQ-learning, approximate GMM methods for five groups. The results showed that dQ-learning (blue) was superior to tQ-learning (red) and approximate GMM (green) in predicting the trajectory. Significant differences are indicated by * \( p < 0.05 \), ** \( p < 0.01 \) and *** \( p < 0.001 \), analyzed by the Kruskal–Wallis H test, with multiple-comparison performed by Dunn’s test. N indicates the total number of neurons. All data are presented as mean ± SEM.
4.3. Cell types influence learning performance

The performance evaluation showed that cell types influence not only the convergence rate but also prediction accuracy. Using dQ-learning with the $P_d \& Nw, \theta$ group resulted in the highest level of accuracy and the fastest convergence rate, indicating that the combination of place cells and interneurons $w, \theta$ may provide more spatial information to predict direction. As pointed out in a previous study, the phase precession not only represents spatial cues about goal location \[^{[69]}\] but also conveys information about ongoing goals or intentions.\[^{[70,71]}\] Furthermore, the phase precession reveals the different phases as the animal runs along a different segment of the maze to different goal locations.\[^{[72]}\] When the reward function of dQ-learning was acting efficiently, the learning progress was approximate to the goal location and actual movement direction. In addition, cQ-learning, based on the adaptive reward function for the goal-navigation task \[^{[73]}\] improves the prediction accuracy but decreases the convergence rate. However, unfortunately the cQ-learning is impossible in practice for real-time decoder due to its large time complexity. The calculation formula for time complexity is $O(|S|^2 \times |A|)$, where $|S|$ is the number of possible states and $|A|$ is the number of possible actions \[^{[73]}\].

In the present study, the time complexity is 720,000 for each time step by $|S| = (30 \times 30 - 15 \times 15) / 1.5 = 300$ and $|A| = 8$ (eight directions). The maximal time complexity of dQ-learning is 120 in Rat 180. The time complexity of cQ-learning was 6000 times longer than that of dQ-learning and made implementation of cQ-learning nearly impossible. Thus, we implemented approximate GMM, which a computationally efficient point-process filter, to compare decoding performance of dQ-learning. Results showed that dQ-learning was superior to approximate GMM in predicting the location of the animal. Previous study found that GMM had a poorer decoding performance when the signals exhibited non-Gaussian distributions\[^{[74]}\]. Although the place fields of place cells usually exhibited Gaussian distributions\[^{[100]}\], those of the interneurons presented the opposite.\[^{[42]}\] Thus, interneurons are unsuitable for use in approximate GMM to predict the locations of the animal. Moreover, the decoding performance of approximate GMM was also affected by the number of neurons because it uses posterior estimate probability distribution to predict the animal location.\[^{[58]}\] Moreover, the decoding performance of approximate GMM was also affected by the number of grid cells because it uses posterior estimate probability distribution to predict the animal location.

5. Conclusions

With good prediction accuracy, we reached the results that the dQ-learning algorithm achieved better learning performance than $\theta$Q-learning and approximate GMM. We also demonstrated that the adaptive reward function and cell types are critical factors for applying dQ-learning method to hippocampal spatial decoding. Our future work will focus on the impact of motion speed in real-world navigation tasks, as a recent study reported that the firing rate of interneurons is strongly correlated with motion speed. The dQ-learning algorithm therefore could be further developed to involve both direction and speed in predicting trajectories for real-world navigation systems.

Acknowledgements

This work is financially supported by Ministry of Science and Technology of Taiwan under Contract numbers of MOST 108-2321-B-010-008-MY2, 108-2814-B-303-010, 108-2636-E-006-010, 109-2636-E-006-010 (Young Scholar Fellowship Program), 107-2221-E-010-021-MY2, and 107-2221-E-010-011. We also are grateful for support from the Headquarters of University Advancement at the National Cheng Kung University, which is sponsored by the Ministry of Education, Taiwan.
References

42. Hangya, B., et al., Complementary spatial firing in place cell-


48. Mamad, O., et al., Extrafield activity shifts the place field center of mass to encode aversive experience. eNeuro, 6(2) (2019).


Appendix A

Circular-linear regression

Phase precession is quantified by a linear model in Figure A1. The circular-linear regression as follows:

\[ \hat{\phi}_j = 2\pi \beta \hat{\theta}_j + \phi_0, \quad j = 1, 2, \ldots, M \]  

(A.1)

where \( \hat{\phi} \), \( \hat{\theta} \), \( \beta \), \( \phi_0 \), and \( M \) are the predicted theta phase, azimuth angle of the polar coordinate, slope, phase offset, and number of spikes of one neuron, respectively. Note that \( \hat{\phi} \) and \( \hat{\theta} \) are both transferred from degree to radian for circular-linear regression. The \( \beta \) and \( \phi_0 \) are estimated by minimizing the circular distance between the measured theta phase \( \phi \) and the predicted theta phase \( \hat{\phi} \), as follows:\[^{[81]}\]

\[ d(\phi_j, \hat{\phi}_j) = 2[1 - \cos(\phi_j - \hat{\phi}_j)] \]  

(A.2)

It has been demonstrated that minimizing \( d \) is equivalent to maximizing \( K \), as follows:\[^{[82]}\]

\[ K = \frac{1}{M} \sum_{j=1}^{M} \sin(\phi_j - 2\pi \beta \hat{\theta}_j) - \frac{1}{M} \sum_{j=1}^{M} \sin(\phi_j - 2\pi \hat{\theta}_j) \]  

(A.3)

The slope \( \beta \) can be estimated as follows:

\[ \beta = \arg\max K \]  

(A.4)

Thus, the phase offset can be estimated as follows:

\[ \phi_0 = \arctan \frac{\sum_{j=1}^{M} \sin(\phi_j - 2\pi \hat{\theta}_j) - \sum_{j=1}^{M} \sin(\phi_j - 2\pi \hat{\theta}_j)}{\sum_{j=1}^{M} \cos(\phi_j - 2\pi \hat{\theta}_j) - \sum_{j=1}^{M} \cos(\phi_j - 2\pi \hat{\theta}_j)} \]  

(A.5)

where \( \arctan^* \) represents the quadrant-specific inverse of the tangent. The theta phase can be predicted by the azimuth angle of the polar coordinate using the linear model.

Furthermore, the circular-circular correlation coefficient between the predicted phase and theta phase is defined as follows:\[^{[82]}\]

\[ \rho_C = \frac{\sum_{j=1}^{M} \sin(\omega_j - \hat{\omega}) \sin(\phi_j - \hat{\phi})}{\sqrt{\sum_{j=1}^{M} \sin^2(\omega_j - \hat{\omega})} \sqrt{\sum_{j=1}^{M} \sin^2(\phi_j - \hat{\phi})}} \]  

(A.6)

where

\[ \omega_j = 2\pi \frac{\bar{\beta}}{\beta} \hat{\theta}_j (\mod 2\pi) \]  

(A.7)

\[ \bar{\phi} = \arctan \frac{\sum_{j=1}^{M} \sin(\phi_j)}{\sum_{j=1}^{M} \cos(\phi_j)} \]  

(A.8)

\[ \bar{\omega} = \arctan \frac{\sum_{j=1}^{M} \sin(\omega_j)}{\sum_{j=1}^{M} \cos(\omega_j)} \]  

(A.9)

\( \bar{\phi} \) and \( \bar{\omega} \) are circular sample mean values.

In Figure A2, we show the distribution of the circular-circular correlation coefficient \( \rho_C \) from three rats in the CW and CCW directions, respectively. The Figure A3 show the Circular linear regression line of all cells from three rats.

Appendix B

The convergence state of the navigation model

The dQ-learning algorithm is a model-free RL method and uses the reward prediction error (RPE) to update the navigation model. The RPE, \( \delta_{RPE} \) was defined as:
\[ \delta_{RPE} = \sum_{t=1}^{T} \left[ \alpha R(t) + \gamma \max_i Q(f(t), a_i) - \psi_i \right] \phi(f_i(t)) \] (B.1)

where \( Q(f(t), a_i) \) and \( i^* \) were obtained from Eq. 3 and Eq. 6, respectively, and \( \gamma = 0.7 \) was the discounting factor. The convergence state of our proposed navigation model was based on value iteration, and the \( \delta_{RPE} \) was used to update the Q-value:

\[ Q(f(t), a_i^*) \leftarrow Q(f(t - 1), a_i^*) + \alpha \cdot \delta_{RPE} \] (B.2)

where \( \alpha = 0.7 \) was the learning rate. Only the Q-value whose action was actually taken at the time step \( t - 1 \) wins the learning chance and was updated at the time step \( t \). The model of Q-learning was chosen when the Q-value was stable and the \( \delta_{RPE} \) tended toward zero.

**Appendix C**

Approximate GMM algorithm

The data include the timestep of the hippocampal neuron (place cell and interneuron) \( s_{x,t} \) and state variable \( X_t \) where \( t \) is the time index. \( X_t \) comprises the position \( x_t \) and velocity \( v_x, v_y \). The approximate GMM is a probabilistic model that assumes that the firing timestamp of the hippocampal neuron \( s_{x,t} \) is generated from a mixture of a finite number of Gaussian distributions with unknown parameters \((u_{s,p}, W_s, \pi_s)\). This algorithm assumes that the posterior distribution of the firing timestamp of the hippocampal neuron \( s_{x,t} \) at each time \( t \) can be approximated by the Gaussian model.

\[ p(X_k|s_{x,t}) \propto \sum_s N(s_{x,t}|u_{s,p}, W_s)\pi_s \] (C.1)

where \( N \) is the Gaussian distribution, \( \pi_s \) is the mixing weight, and \((u_{s,p}, W_s)\) is the mean vector and covariance matrix of the \( s^p \) mixture component. We assume that there are \( S \) mixtures in total. \( S \) is determined based on the number of recorded hippocampal neurons. The parameters of approximate GMM \((u_{s,p}, W_s, \pi_s)\) are determined by the expectation-maximization (EM) algorithm along with the Bayesian information criterion (BIC). The approximate GMM algorithm is tuned in conformity with the following four steps:

1. Run the EM algorithm to identify a GMM with a conservatively large number of mixture components \( S \).
2. Calculate the expectation of the log-likelihood, which is evaluated using the current estimate for the parameters \((u_{s,p}, W_s, \pi_s)\) at each pair of the \( S^n \) mixture component. Remove the pair with the lowest expectation.
3. Replace the removed pair with a new mixture component that maximizes the cost function. Estimate the mean, covariance, and mixing weight of the new mixture component \((u_{s,p}, W_s, \pi_s)\) using the EM algorithm.
4. Determine the BIC of the reduced model. The BIC of the original model with \( S \) mixture components is defined as follows:

\[ \text{BIC}_S = -2 \sum_p \log(\sum_s \pi_s L(x^p; u_{s,p}, W_s)) + \ln(P)D_g \] (C.2)

\[ D_g = S \left( \frac{d^2-d}{2} + 2d + 1 \right) - 1 \] (C.3)

where \( x^p \) is the \( p^h \) sample data from the proposal distribution, \( P \) is the number of sample data, \( d \) is the...
dimension of $x^p$, and $L(x^p; u_s, W_s)$ is the likelihood of $x^p$ from a multivariate normal distribution with $u_s$ and $W_s$ parameters. Select one model with the lowest BIC between the reduced model $(u_E, W_E, \pi_E)$ and the original model $(u_s, W_s, \pi_s)$. Repeat from Step 2) until the BIC fails to reduce.

Finally, a GMM consisting of a minimum number of mixture components can be found. The Figure C1 shows the distribution of the BIC at each number of components, and the optimum number of components of the GMM with the lowest BIC in each group from three rats as shown in Table C1.

Table C1. The optimum number of components selected via minimum BIC for the approximate GMM.

<table>
<thead>
<tr>
<th>Subject</th>
<th>Rat 180</th>
<th>Rat 187</th>
<th>Rat 213</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P$</td>
<td>CW</td>
<td>CCW</td>
<td>CW</td>
</tr>
<tr>
<td>$IN$</td>
<td>4</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>$P$ &amp; $IN_w, \theta$</td>
<td>7</td>
<td>6</td>
<td>5</td>
</tr>
<tr>
<td>$P$ &amp; $IN_w, \theta$</td>
<td>6</td>
<td>6</td>
<td>4</td>
</tr>
<tr>
<td>$P$ &amp; $IN$</td>
<td>7</td>
<td>7</td>
<td>5</td>
</tr>
</tbody>
</table>