Dynamic causal modeling of hippocampal activity measured via mesoscopic voltage-sensitive dye imaging

Jiyoung Kang, Kyesam Jung, Jinseok Eo, Junho Son, Hae-Jeong Park

PII: S1053-8119(20)30242-1
DOI: https://doi.org/10.1016/j.neuroimage.2020.116755
Reference: YNIMG 116755

To appear in: NeuroImage

Received Date: 8 October 2019
Revised Date: 10 March 2020
Accepted Date: 14 March 2020

Please cite this article as: Kang, J., Jung, K., Eo, J., Son, J., Park, H.-J., Dynamic causal modeling of hippocampal activity measured via mesoscopic voltage-sensitive dye imaging, NeuroImage (2020), doi: https://doi.org/10.1016/j.neuroimage.2020.116755.

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2020 Published by Elsevier Inc.
Jiyoung Kang: Conceptualization, Methodology, Software, Formal analysis, Visualization, Writing- Original Draft, Reviewing and Editing. Kyesam Jung: Methodology, Software. Jinseok Eo: Software. Junho Son: Software. Hae-Jeong Park: Conceptualization, Methodology, Visualization, Writing- Reviewing and Editing, Supervision, Funding acquisition
Dynamic causal modeling of hippocampal activity measured via mesoscopic voltage-sensitive dye imaging

Jiyoung Kang\textsuperscript{1,2}, Kyesam Jung\textsuperscript{1,2}, Jinseok Eo\textsuperscript{3}, Junho Son\textsuperscript{3}, and Hae-Jeong Park\textsuperscript{1,2,3,4,}\*

\textsuperscript{1}Center for Systems and Translational Brain Sciences, Institute of Human Complexity and Systems Science, Yonsei University, Seoul, Republic of Korea
\textsuperscript{2}Department of Nuclear Medicine, Yonsei University College of Medicine, Seoul, Republic of Korea
\textsuperscript{3}BK21 PLUS Project for Medical Science, Department of Radiology, Department of Psychiatry, Yonsei University College of Medicine, Seoul, Republic of Korea
\textsuperscript{4}Department of Cognitive Science, Yonsei University, Seoul, Republic of Korea,

*Corresponding author:
Hae-Jeong Park, PhD
Professor, Department of Nuclear Medicine, Radiology, Psychiatry, Yonsei University College of Medicine, Department of Cognitive Science, Yonsei University,
Center for Systems Neuroscience, Institute of Human Complexity and Systems Science,
50 Yonsei-ro, Sinchon-dong, Seodaemun-gu, Seoul, Republic of Korea 120-752,
Tel.: 82-2-2228-2363; Fax: 82-2-393-3035
Email: parkhj@yonsei.ac.kr
The aim of this paper is to present a dynamic causal modeling (DCM) framework for hippocampal activity measured via voltage-sensitive dye imaging (VSDI). We propose a DCM model of the hippocampus that summarizes interactions between the hilus, CA3 and CA1 regions. The activity of each region is governed via a neuronal mass model with two inhibitory and one/two excitatory neuronal populations, which can be linked to measurement VSDI by scaling neuronal activity. To optimize the model structure for the hippocampus, we propose two Bayesian schemes: Bayesian hyperparameter optimization to estimate the unknown electrophysiological properties necessary for constructing a mesoscopic hippocampus model; and Bayesian model reduction to determine the parameterization of neural properties, and to test and include potential connections (morphologically inferred without direct evidence yet) in the model by evaluating group-level model evidence. The proposed method was applied to model spatiotemporal patterns of accumulative responses to consecutive stimuli in separate groups of wild-type mice and epileptic aristaless-related homeobox gene (Arx) conditional knock-out mutant mice (Arx<sup>-/-</sup>;Dlx5/6<sup>Cre-ires-GFP</sup>) in order to identify group differences in the effective connectivity within the hippocampus. The causal role of each group-differing connectivity in generating mutant-like responses was further tested. The group-level analysis identified altered intra- and inter-regional effective connectivity, some of which are crucial for explaining mutant-like responses. The modelling results for the hippocampal activity suggest the plausibility of the proposed mesoscopic hippocampus model and the usefulness of utilizing the Bayesian framework for model construction in the mesoscale modeling of neural interactions using DCM.

**Keywords**

Neural population analysis, computational modeling, effective connectivity, dynamic causal modeling, voltage sensitive dye image, hippocampus
INTRODUCTION

Dynamic causal modeling (DCM) (Friston et al., 2007; Friston et al., 2003) has been used to model effective connectivity (directional, asymmetrical influences of one node over another) among multiple neural populations using various levels of neuroimaging data, such as electroencephalograms, functional MRIs, and local field potentials (Friston et al., 2017; Kiebel et al., 2008; Moran et al., 2013). Most of these studies model macroscopic interregional neural circuitry, which are directly or indirectly associated with local field potentials of neural populations. However, studies that model mesoscopic neural interactions reflected in the membrane potentials by using DCM are rarely found despite the wide usage of voltage-sensitive dye imaging (VSDI) or calcium imaging (indirectly) in measuring membrane potentials. Only a few DCM studies for calcium imaging have recently been introduced (e.g., Jung et al., 2019; Rosch et al., 2018) but not for VSDI. In terms of modelling a neural circuitry using DCM in the mesoscopic level, VSDI is highly advantageous since it measures membrane potentials of distributed neural cells or neural populations with high temporal (1–2 ms) and spatial (20–50 μm) resolution (Chemla and Chavane, 2010b).

The essence of DCM, particularly regarding mesoscopic modelling of a neural circuitry, lies in the ability to construct a biologically plausible but simple neural state model. To make a biologically plausible model, researchers often refer to previous literature for model architecture and/or neural properties. For example, for the hippocampus, one may refer to public databases of the hippocampus to employ electrophysiological properties such as resting membrane potential, action potential threshold and time decay constant of neural populations. Such electrophysiological properties work as reference values (constants) within a DCM model, based on which model parameters (e.g., connectivity parameters) are estimated during the model inversion process. However, some reference values of neural properties that are necessary in the model equation are not yet available, and some depend on the experimental condition. Thus, reference values, particularly for neural properties without previous references, should be determined before the application of DCM to simplify model inversion. So far, adequate methods to determine not-yet-available reference values for a model for DCM have not been researched sufficiently. In this study, we propose a method to determine unknown (unavailable) reference values using a Bayesian optimization technique, which is widely used in the fields of machine learning (for review, see Shahriari et al. (2016)). Since unknown reference values should be optimized before parameter estimation in DCM, we will refer to the unknown reference values as hyperparameters in this study.
Another requirement of computational modeling using DCM is to make a model structure as simple as possible, but not too simple (optimizing a model structure within the constraint of biological plausibility). This is especially critical for mesoscopic modeling with multiple parameters for neural properties, which inevitably induce high complexity. This optimization of a model structure should be conducted in the model construction stage, by removing redundant or question-irrelevant parameters from the model. For example, we need to determine whether or not to assign modulation parameters for reference values (either directly obtained from previous literature or derived from hyperparameter optimization) to account for inter-individual (or inter-group) variability at the cost of increased model complexity in DCM. We also need to simplify the model architecture based on experimental data. For example, when modeling the hippocampus, we may determine whether to include or ignore some connections among neural populations which are presumed to exist but without direct evidence (Wheeler et al., 2015). All these issues are not trivial, but they need to be resolved in order to successfully apply mesoscopic DCM to real experimental data analysis.

The first goal of the current study is to propose practical schemes for implementing mesoscopic modeling of neural interactions using a DCM framework. First, we introduce Bayesian optimization of hyperparameters in high-dimensional space to select unknown reference values for DCM. Second, as a step for simplifying model structure, we adopt a model selection scheme to determine (i) whether to use certain reference values as constants or as variables to be estimated in DCM and (ii) whether or not to include specific connections (edges) in the model, by comparing group-level model evidence. Third, we propose a method to evaluate the causal role of statistically significant parameters in generating target responses. The computational modeling using DCM often ends with the group-level test for model parameters. However, the statistical difference of a parameter between groups does not necessarily indicate the causal role of the parameter (e.g., effective connectivity) in inducing signal differences between groups. For this, we propose a method to evaluate each parameter with respect to causality.

Upon establishing technical schemes for optimizing model structure, the second goal of the current paper is to propose a DCM model for hippocampal responses to stimuli measured using VSDI. The hippocampus is a complex center for learning, emotion, memory (Roy et al., 2017), and spatial navigation (Eichenbaum and Cohen, 2014). The hippocampus is also a pathologically important organ in various brain diseases such as Alzheimer's disease (Braak and Braak, 1991; Masurkar, 2018) and temporal lobe epilepsy (Ang et al., 2006; Krook-Magnuson et al., 2013; Laurent et al., 2015; Sanjay et al., 2015). Because of the critical role of the hippocampus, the functional circuitry of the normal and abnormal hippocampus has been a compelling target for computational modeling for diverse purposes particularly at
micro-scales (e.g., Cutsuridis et al., 2018). Most computational modeling studies of the hippocampus have focused on intra-regional connectivity at a specific region of the hippocampus (e.g., Bezaire et al., 2016; Sanjay et al., 2015), but not on inter-regional connectivity among the hippocampus regions. In the current study, we demonstrate a computational modelling of both intra- and inter-regional interactions within the hippocampus that is responding to an external stimulus, using a DCM framework.

So far, many DCM studies have been conducted to explain interactions among cortical regions or between cortical and subcortical regions, based on a convolution-based framework (Bastos et al., 2012; Jansen and Rit, 1995; Moran et al., 2013). However, for the hippocampus model, since the architecture of neural compositions differ from that of cortical columns, we did not reuse conventional neural state models designed for cortical columns. Instead, we propose a DCM model for the hippocampus, with neurobiological structure and properties that are specific to subregions of the hippocampus as found in previous experimental studies. In particular, the proposed model for the hippocampus consists of one (or two) excitatory and two inhibitory neural populations for a column since the hippocampus consists of relatively homogeneous excitatory but diverse inhibitory neurons (Bezaire et al., 2016; Kullmann, 2011; Sanjay et al., 2015; Wendling et al., 2002). This composition differs from that of the Jansen-Rit model (two excitatory and “one inhibitory”) (Jansen and Rit, 1995; Moran et al., 2013) and canonical microcircuit models (three excitatory and “one inhibitory”) (Bastos et al., 2012; Moran et al., 2013) designed for the cortical column.

Using the proposed hippocampus model, we analyzed the effective connectivity in two different groups of mice, available from an open VSDI dataset (Bourgeois et al., 2014). The dataset contains stimulation-locked VSDI of hippocampal brain slices of wild-type and epileptic aristaless-related homeobox gene (Arx) conditional knock-out mutant mice (Arx\(^{+/−}\); Dlx5/6\(^{CRE-IRES-GFP}\)) (Marsh et al., 2009). According to Bourgeois et al. (2014), not only the signal amplitude but also the spatiotemporal activity evoked by consecutive 100-ms interval stimuli differs between that of the wild-type mice and the that of the epileptic mutant mice. Specifically, for the first stimulus, hyperpolarization of the cornu ammonis 1 (CA1) region was observed in both wild-type and mutant mice. However, for subsequent stimuli, hyperpolarization of the CA1 region was reduced in wild-type mice while it was maintained in mutant mice (Figure 1). Furthermore, accumulative signals (stronger responses for consecutive stimuli) particularly in the hilus region were only observed in wild-type mice (Figure 1).

To explain the differing pattern of accumulative responses for consecutive stimuli across regions and groups (regardless of scale factors), we focused on effective connectivity among neural populations both within and between regions of the hippocampus, rather than local
neural properties of the hippocampus. However, not much is known about the inter-regional effective connectivity in response to an external stimulus, for both healthy and altered hippocampus. This may partly be due to the fact that the effective connectivity among neural populations cannot be directly assessed by simply measuring the activity of each neural population, but calls for computational modeling, like DCM. Regarding the altered responses to external stimuli in the conditional knock-out mutant mice (Arx<sup>+/−</sup>;Dlx5/6<sup>CRE-IRESCre</sup>-GFP), a previous study reported regional alterations in gene expression, e.g., reduced Arx expression in the DG and increased Arx expression in the CA1 of the hippocampus in young mutant mice and only slightly reduced Arx expression in the CA3 in adult mutant mice (Marsh et al., 2016). However, how these regional alterations of Arx expression (and possibly neural composition) by mutation in the childhood affect the intra- and inter-regional connectivity of the adult mice has not been explored. By using computational modelling with appropriate effective connectivity, we expected to find a potential connectivity-based mechanism that produces different patterns of accumulative waveforms in response to consecutive stimuli, and to identify the inter-regional effective connectivity responsible for altered responses in the hippocampus of the mutant mice using DCM for VSDI.

The current paper is composed of four parts: 1) introduction of data sets and construction of neural state and observation models using VSDI-DCM for the hippocampus; 2) Bayesian hyperparameter optimization methods for DCM; 3) group-level comparison of wild-type and mutant mice; and 4) perturbation analysis to evaluate the causal role of each connection in eliciting the abnormal behaviors in the mutant mice.

MATERIALS and METHODS

VSDI dataset

The present study utilized an open VSDI dataset of the hippocampal brain slices of wild-type mice and epileptic Arx conditional knock-out mice (Arx<sup>+/−</sup>;Dlx5/6<sup>CIG</sup> mouse) after electrical stimulation at the temporoammonic (TA) pathway (Bourgeois et al., 2014).
Figure 1. VSDI observed at the hippocampus. (A) Three regions in the hippocampus (the hilus, CA3 and CA1 regions), (B) group-average (N = 5) VSDI signals along the stratum radiatum of the hippocampus for the wild-type mice and mutant mice, and (C) group-average VSDI signals at the three hippocampal regions for the wild-type mice (upper part) and mutant mice (lower part) are presented. The arrow in (A) indicates the location of stimulation (electrode). Raster plots for VSDI signals from the 44 columnar regions in (B) is visualized using VSDI MATLAB toolbox (Bourgeois et al., 2014).

The open dataset includes VSDI signals of 44 segments (segmented by 0.1 mm) along the stratum radiatum (Figure 1A and 1B) of the hippocampal brain slices. Bourgeois et al. (2014) reported that knock-out mice exhibited frequent spontaneous seizures and behavioral deficits, showing symptoms similar to humans with mutations in the ARX gene. Details of experiment are available at Bourgeois et al. (2014).

The VSDI signals in the dataset were recorded under four repetitive instantaneous (0.1 ms) electric stimuli inserted near the TA pathway with 100 ms intervals. For the first stimulus, hyperpolarization after the stimulation was observed in the CA1 region, but this inhibition was reduced for the following consecutive stimuli. This experiment is different from other experiments (e.g., Ang et al., 2006), in which the CA3-CA1 connection was cut to block direct effects from CA3 to CA1 and thus to eliminate effect of stimuli from the perforant path. In contrast, the unblocked stimuli in the current dataset may affect all the DG, CA3 and CA1 regions. Therefore, the early coactivation at multiple regions in response to the stimuli could have originated not only from the connection between the CA3 and CA1 regions but also from direct electrical conductivity.

To simplify the computational modeling of hippocampal responses to external stimuli, we extracted VSDI signals from three regions of interest (ROIs) in the label map of Bourgeois et al. (2014), which correspond to the centers of the hilus, CA3 and CA1 regions. Among ten
and eight slices (each slice refers to an individual in this study) provided for wild-type and
mutant mice, respectively, we only utilized the five best slices for each group (chosen by
visual screening) for the subsequent analysis, and excluded slices with very weak and noisy
responses. We normalized the extracted signals for each group using the maximal sum of
entire VSDI signals of the slices, and applied a median filter with a 10 ms window, as used in
Bourgeois et al. (2014).

Basics of the convolution-based neural state dynamic model

The VSDI-DCM framework for the hippocampus consists of two parts of modeling; a
neural state dynamic model for the hippocampus and an observation model for VSDI. For the
neural state dynamic model in the hippocampus, we used a convolution-based model (Jansen
and Rit, 1995; Moran et al., 2013) composed of two consecutive transform functions $S(V)$ and
$h(t)$.

The sigmoidal (activation) function $S(V_j)$ of a neural population $j$ describes a
transformation of the average membrane potential $V_j$ to the average firing rate of action
potentials, denoted by,

$$S(V_j) = \frac{f_j}{1 + e^{-R_j(V_j - V_{0,j})}}.$$  \hspace{1cm} (Eq. 1)

with parameters of a maximal firing rate $f_j$, a post-synaptic potential (PSP) reference $V_{0,j}$ and a
slope $R_j$ of the sigmoid function. $V_{0,j}$ is the PSP that achieves a $50\%$ firing rate of a neural
population $j$ (Jansen and Rit, 1995).

The synaptic kernel $h_i(t)$ is a transformation function from an average firing rate coming
to a neuronal population $i$ to an average post-synaptic membrane potential at the neural
population $i$ (which can either be inhibitory or excitatory) with a time constant of synaptic
kernel $\tau_i$ and a maximal PSP $H_i$.

$$h_i(t) = \begin{cases} \frac{H_i}{\tau_i} t \cdot \exp \left( -\frac{t}{\tau_i} \right), & \text{for } t \geq 0 \\ 0, & \text{for } t < 0 \end{cases}.$$ \hspace{1cm} (Eq. 2)

Thus, the average membrane potential $(V_i)$ of a neural population $i$, evoked by an average
membrane potential $(V_j)$ of a neural population $j$, can be described as follows.

$$v_i(t) = h_i(t) \otimes S(V_j) = \int_{-\infty}^{t} h_i(t - \tau) S(V_j(\tau)) d\tau.$$ \hspace{1cm} (Eq. 3)

If we extend Eq. 3 to multiple neural populations coming to a neural population $i$ with
different connectivity $A_{ij}$, the convolution-based neural state model can be written by using
the following ordinary differential equations:
\[
\dot{V}_i = I_i, \quad \text{(Eq. 4)}
\]
\[
\dot{I}_i = \frac{H_i}{\tau_i} \left( \sum_{j} A_{ij} S(V_j) + C_i u_i \right) - \frac{2I_i}{\tau_i} - \frac{V_i}{\tau_i}^2. \quad \text{(Eq. 5)}
\]

Here, \(V_i\) and \(I_i\) indicate the membrane potential and cross membrane current of a neural population, respectively. The effective connectivity from a neural population \(j\) to a neural population \(i\) is denoted by \(A_{ij}\). The external input for the neural populations \(u_i(t)\) is multiplied by input modulation parameter \(C_i\). Of note, since the current experiment does not have experimental conditions as within-subject factors, no modulation input (corresponding to B matrix in the conventional DCM for event related potential (ERP)) was assigned to approximate the VSDI time series in the current study.

Neural state dynamic model for the hippocampal activity

The hippocampus has principal cell bodies located in a layer with dendritic trees and axonal projections in the lamina and is composed of populations of excitatory (principal cells) and inhibitory neurons (interneurons) (Booker and Vida, 2018). The principal cells of the hippocampus, among diverse neural cell types, are pyramidal cells of the CA1 and CA3 regions and granule cells of the dentate gyrus (DG) (Hamilton et al., 2017), and are mostly excitatory. Although the numbers of excitatory neurons are larger than those of inhibitory neurons, excitatory neurons are relatively homogenous (Booker and Vida, 2018). Meanwhile, inhibitory neurons exhibit more diverse characteristics (e.g., different firing patterns, calcium-binding proteins, and axonal targets) (Freund and Buzsáki, 1996; McBain and Fisahn, 2001) and play critical roles in controlling the hippocampus during learning and memory or in inducing pathological states (e.g., epilepsy, Alzheimer’s disease) (Allen and Monyer, 2015; Krook-Magnuson et al., 2013). Therefore, we constructed a neural state model for the hippocampus that consists of one (or two) excitatory and two inhibitory neural populations for a column, based on a convolution-based framework (Bastos et al., 2012; Jansen and Rit, 1995; Moran et al., 2013).

To construct a model for VSDI of the hippocampal slices of the wild-type and mutant mice, we first chose three representative regions of the hippocampus (the hilus, the CA3 and CA1 regions), and modelled the interactions in those regions with ten neural populations. The ten populations are 1) the granule, mossy, DG Basket and HIPROM in the hilus; 2) CA3c Pyramidal, CA3 Basket and CA3 O-LM in the CA3 region; and 3) CA1 Pyramidal, CA1 Basket, and CA1 recurrent O-LM cells (Figure 2). They were chosen according to the locations of their dendrites (obtained from a morphology matrix in the Hippocampome open database (Wheeler et al., 2015)—For the hilus region, two excitatory neural populations
(granule and mossy cells) and two inhibitory neural populations (DG Basket and HIPROM
cells) were assigned. Both the CA3 and CA1 regions were again composed of one excitatory
neural population (pyramidal cell) and two inhibitory neural populations (Basket and O-LM
cells).

Figure 2. The neural state model topology used for the hippocampus. Ten neural populations
(dotted circle: latent (hidden) neural populations) and all of their possible connections (solid
line: known connection, dotted line: potential connection) are assumed. Excitatory and
inhibitory connections are colored with red and blue, respectively. The neural populations are;
E11 (granule), E12 (mossy), I11 (DG Basket), I12 (HIPROM), E21 (CA3c Pyramidal), I21
(CA3 Basket), I22 (CA3 O-LM), E31 (CA1 Pyramidal), I31 (CA1 Basket), and I32 (CA1
recurrent O-LM). External electric shocks were presumed to affect all regions of the stratum
radiatum and dentate gyrus. Thus, the external electric stimulus \( u \) is configured to affect E11,
E12, I11, E21, I21, E31, and I31, directly. Considering the distance from the stimulus, we did
not set direct input to two neural populations in the stratum oriens (I22 and I32).

The electrophysiological properties such as maximal firing rate \( f_i \), decay time constant
\( \tau_i \), thresholds for action potential \( V_{th,i} \) and resting membrane potentials \( V_{rest,i} \) were
employed from the open database (Table 1). However, reference values for the PSP \( V_{0,j} \) and
slope \( R_j \) of the sigmoid function in Eq. 1 are not available in the database. Thus, reference
values for \( V_{0,j} \) and \( R_j \) for each neural population need to be estimated.

Previous studies (Jansen and Rit, 1995; Moran et al., 2013) have assigned the PSP \( V_0 \)
(initiating at a 50% firing-rate) and slope \( R_j \) as fixed constants regardless of neural populations.
However, we assigned the reference PSP \( V_{0,j} \) variable for each neural population \( j \) in order to
allow for diverse neural behaviors. To reduce the number of model parameters, we assumed
that the PSP reference \( V_{o,j} \) at a neural population \( j \) can roughly be approximated from a
voltage shift from its action potential threshold \( V_{th,j} \), as shown below.

\[
V_{o,j} = V_{th,j} + V_{offset}, \quad V_{offset} > 0 \quad \text{(Eq. 6)}
\]

We also assumed a homogeneous slope across all neural populations (i.e., all \( R_j \) are assumed
to have the same value \( R \)). Previously, a steep sloped sigmoid function was used, such as \( R = 0.67 \text{ mV}^{-1} \) in Moran et al. (2013). However, by parameterizing \( R \), we expected to find a lower \( R \) value (a slower slope) to extend the range of monotonic increases in the sigmoid function
for each individual. Therefore, the parameters \( R \) and \( V_{offset} \) are to be estimated to explain a
wider response range and diversity of each neural population in the sigmoid activation
function of each individual.

Although neural populations in the stratum oriens and dentate are presumed to be
involved in the processing of the external stimulus, the VSDI signals for these regions are
unavailable in the public dataset. Therefore, we employed such neural populations (i.e., O-
LM neural populations) as latent (hidden) in the model. Those latent neural populations
constitute the dynamics of neural states according to Eq. 5 but their direct contributions to the
observed VSDI signals were set to zero in the observation model.

The Hippocampome open database provides not only known (identified in previous
literatures) but also potential connections (morphologically inferred without direct evidences
yet) which were presumed to exist based on Peter’s rule (Rees et al., 2017). According to
Peter’s rule, axon-dendrite colocations are considered as proxies of potential connections
(Rees et al., 2017). Thus, in the current hippocampus model, we constructed a full model
composed of both types of connections, comprising a total of 41 connections (17 known and
24 potential connections) (Figure 2). Then, we evaluated and compared reduced models
nested in the full model using Bayesian model reduction to test the need for potential
connections in the proposed model. The details will be explained in a later section.

To account for potential spatiotemporal delays to diverse neurons in a population during
the propagation of electric shock as an external input, we used a Gaussian-convolved
waveform \( u(t) \) at four onset times,

\[
u(t) = \sum_{i=1}^{4} \exp\left( -\frac{(t-t_i)^2}{\tau} \right), \quad t_i = 70, 170, 270, 370 \text{ ms},
\]

with a narrow Gaussian kernel size of \( \tau = 0.2 \). The input was assigned to all neural
populations within the radiatum and hilus, which are close to the location of the electric
stimuli (see Figure 2).
Observation model for VSDI signal

A linear relation between a VSDI signal and membrane potential has been reported in previous studies (Berger et al., 2007; Chemla and Chavane, 2010a). Accordingly, we employed a linear observation model for the VSDI data. The VSDI signal $v_r$ at a region $r$ composed of $n$ neural populations is described as follows.

$$v_r = \alpha_r \sum_{i=1}^{n} \rho_{ri} V_{ri}, \quad \text{(Eq. 7)}$$

$$\sum_{i=1}^{n} \rho_{ri} = 1, \quad \text{for all populations in the region } r, \quad \text{(Eq. 8)}$$

in which $\alpha_r$ represents a scaling factor of a signal at a region $r$ to fit the VSDI signal amplitude. $V_{ri}$ and $\rho_{ri}$ indicate a membrane potential at a population $i$ and its contribution ratio to the VSDI signal in region $r$. In the observation model for VSDI, $\alpha_r$ and $\rho_{ri}$ were set as variables to be estimated. The scaling factor $\alpha_r$ reflects combinations of the regional neural sensitivity, optical factors and other unspecified factors common across neural populations in region $r$. Among cell-type dependent factors, the contribution ($\rho_{ri}$) of each neural population to a VSDI signal in region $r$ is assumed to be primarily determined by the neural cell density of each population. In the hippocampus, the density of excitatory neurons is known to be much higher than those of inhibitory neurons (Keller et al., 2018). Following this as a principle, we assigned high and low contributions to excitatory and inhibitory neural populations, making the search space for excitatory neural population between 0.5 and 1.0 (i.e., 50% ~ 100%) and inhibitory population between 0.1 and 0.5 in the hyperparameter optimization. These values were sequentially optimized in the hyperparameter optimization, and final results are listed in Table 1.

In summary, model parameters in both neural and observation models include parameters for neural properties, interactions (or synaptic connectivity) among neural populations, and contribution ratios for membrane potentials in the observation model. Reference values for the neural properties of all types of neural cells used in this study are listed in Table 1.

Note that polarity signs for all neural properties were determined beforehand, i.e., positive and negative signs were used to represent the effects from excitatory and inhibitory populations, respectively. Thus, parameters for the neural properties are confined to positive values. The positivity of parameters can be achieved by utilizing either exponential transformation or absolute transformation. The effective connectivity $A_{ij}$ was formulated by
absolute transformation of a parameter $\theta$, i.e., $A_{ij} = \text{abs}(\theta)$, which follows the normal distribution $N(A_{\text{intra/inter}}, 1/32)$, with $A_{\text{intra/inter}}$ as a reference value for intra- or inter-regional connectivity. Using absolute transformation is advantageous in making the effect size zero (or no connection) in the effective connectivity, which is not trivial in the exponential transformation. Other parameters were given by the multiplication of reference values $C_0$ and exponentiated parameters $\theta$, e.g., $C_i = C_0 \exp(\theta_i)$ (Table 2).

Parameters to be estimated are listed in Table 2. Despite simplification of the neural state model for the hippocampus, there are still too many unknown reference values and parameters to be estimated. Therefore, the next step is to confine reference values within biologically plausible ranges and to reduce model parameters in constructing a simplified realistic model. Figure 3 summarizes the model construction procedure for VSDI-DCM for the hippocampus conducted in the current study.

Table 1. Reference values used in the present hippocampus model

| ID  | Name   | Region | $V_{\text{rest}}$ [mV] | $\tau_i$ [ms] | $f_i$ [Hz] | $V_{i\text{hi}}$ [mV] | Reference                                      | $H_{\text{inh/ex}}$ [mV] | $\rho_{ri}$ |
|-----|--------|--------|-------------------------|---------------|-----------|-------------------|-----------------------------------------------|-------------------------|------------|
| E11 | Granule| DG     | -75.5                   | 7.48          | 72.0      | -46.7             | Patel et al. (2004) Lübke et al. (1998)       | 55                      | 0.42       |
| E12 | Mossy  | DG     | -62.0                   | 41.0          | 50.0      | -38.3             | Lübke et al. (1998)                           | 55                      | 0.38       |
| I11 | Basket | DG     | -62.0                   | 10.0          | 230.0     | -43.7             | Lübke et al. (1998)                           | 190                     | 0.09       |
| I12 | HIPROM | DG     | -65.0                   | 35.0          | 69.0      | -38.8             | Lübke et al. (1998)                           | 190                     | 0.11       |
| E21 | Pyramidal | CA3 | -62.0                   | 16.0          | 40.0      | -47.0             | Buckmaster et al. (1993) Spruston et al. (1997) | 55                      | 0.86       |
| I21 | Basket | CA3    | -58.5                   | 11.2          | 33.6      | -32.9             | Gloveli et al. (2005)                         | 190                     | 0.14       |
| I22 | O-LM   | CA3    | -60.0                   | 33.3          | 182       | -33.5             | Gloveli et al. (2005)                         | 190                     | 0          |
| E31 | Pyramidal | CA1 | -62.8                   | 26.1          | 32.0      | -50.0             | Routh et al. (2009) Fernandez et al. (2011)   | 55                      | 0.81       |
| ID | Neuron Type | Region | $V_{\text{rest}}$ | $\tau_i$ | $f_i$max | $V_{\text{th},i}$ | Reference | $\rho_i$ |
|----|-------------|--------|-----------------|--------|---------|-------------|-----------|--------|
| I31 | Basket CA1 | -57.0 | 13.0 | 70.0 | -32.0 | Triccoire et al. (2011) | 190 | 0.19 |
| I32* | Recurrent O-LM CA1 | -75.0 | 12.8 | 150.0 | -50.0 | Ali and Thomson (1998) | 190 | 0 |

Abbreviations: resting membrane potential $V_{\text{rest}}$, time constant of the synaptic kernel $\tau_i$, maximal firing rate $f_i$, threshold for action potential $V_{\text{th},i}$, reference values of the maximal PSP for inhibitory and excitatory neural populations $H_{\text{inh/exc}}$, and contribution ratio of the membrane potential at a neural population $i$ to the VSDI signal at region $r$, $\rho_i$.

1. The first latter of ID represents excitatory I and inhibitory (I).
2. To obtain optimal neural population dynamics, the threshold for the action potential $V_{\text{th},i}$ of each neural population was used to determine the PSP at the half-maximal firing rate ($V_{0,i}$) of the activation function.
3. $H_{\text{inh/exc}}$ are the reference values of $H_i$. $H_i$ indicates the maximum PSP for the $i$-th (excitatory or inhibitory) neural population, $H_i = H_{\text{inh/exc}} \exp(\theta_i)$, $\theta_i \sim N(0, 1/32)$.
4. Latent neural populations that indirectly affect the observed signals ($\rho_i = 0$).
Table 2. Parameters and their priors

| Parameter  | Description                                                                 | Value               | Optimization¹ |
|------------|-----------------------------------------------------------------------------|---------------------|---------------|
| $V_{\text{rest},i}$ | Resting membrane potential, used as an initial value of $V_i(t)$ | Table 1 Fixed      |               |
| $R$        | Slope of a sigmoid function                                                 | 0.21 hyperOpt      |               |
| $V_{\text{offset}}$ | Offset from each membrane potential for actional potential in the sigmoid function | 7.04 hyperOpt |               |
| $f_i$      | Maximal firing rate                                                         | Table 1 Fixed      |               |
| $\tau_i$  | Time constant of the synaptic kernel                                        | Table 1 Fixed      |               |
| $H_{\text{exc}}$² | Maximal PSP reference for excitatory neural populations | 55 hyperOpt |               |
| $H_{\text{inh}}$² | Maximal PSP reference for inhibitory neural populations        | 190 hyperOpt       |               |
| $A_{ij}$³,⁴ | Relative effective connectivity                                              | $A_{\text{intra}}=0.79$ | hyperOpt & DCMopt |
| $C_i$⁵     | Strength of input external stimuli to an unhidden neural population $i$    | Estimated          | DCMopt        |
| $\alpha_r$ | Scaling factor for activity of neural populations in region $r$, in the wild-type and (mutant) groups | 1.30 (0.50) hyperOpt |               |
| $\rho_{ri}$ | Ratio of neural population $i$ in region $r$                               | Table 1 hyperOpt   |               |

¹ hyperOpt indicates Bayesian hyperparameter optimization for reference values. DCMopt (noted as “Estimated”) indicates optimization within the DCM framework.  
² $H_{\text{inh/exc}}$ were determined by Bayesian hyperparameter optimization, and were used as a reference value of maximal PSP $H_i$ of neural population $i$, i.e., $H_i = H_{\text{inh/exc}} \exp(\theta_i)$, $\theta_i \sim N(0, 1/32)$.  
³ $A_{\text{intra}}$ and $A_{\text{inter}}$ represent expectations in the Gaussian distribution (priors) of the intra- and inter-regional effective connectivity. Since the self-recurrent connectivity of the CA1 region is known to be weaker than that of the CA3 region, we assigned $A_{\text{inter}}$ instead of $A_{\text{intra}}$ for the self-recurrent effective connectivity of a pyramidal population in the CA1 region.  
⁴ $A_{ij} = \text{abs} (\theta_{ij})$, $\theta_{ij} \sim N(A_{\text{intra/inter}}, 1/32)$  
⁵ $C_i = C_0 \exp(\theta_i)$, $\theta_i \sim N(0, 1/32)$
Figure 3. The procedure and variables for VSDI-DCM for the hippocampus. To construct a biologically plausible model, we referred literatures and public database (Hippocampome) to determine reference values used in the model as constants (A). Parameters which are not available from the database were optimized by performing Bayesian hyperparameter optimization (B and C). The fixed (same across individuals) and unfixed (allowing individual variances) usage of the maximal post-synaptic potential $H_i$ for each neural population $i$ were determined by using the Bayesian model reduction scheme (D). Group-level Bayesian model comparisons were performed to find the best model of connection combinations (E). For the best model, we performed group comparison analysis to extract group differences in effective connectivity (F). Finally, using perturbation analysis, the causal effect of each effective connectivity in causing abnormal behavior in the mutant mice was evaluated (G). The right upper panel summarizes the reference values, hyperparameters and parameters. In the DCM (using the variational expectation-maximization, EM), the initially chosen reference values and the optimized hyperparameters (unknown reference values searched using Bayesian hyperparameter optimization) were used as constants in the DCM model. The effective connectivity $A_{ij}$, input modulation scale $C_i$, and maximal post-synaptic potential $H_i$ across neural populations, were parameters estimated in the DCM. It should be noted that variational EM-DCM also estimates hyperparameter $\lambda$ for error and parameter covariances, which differ from hyperparameters of unknown reference values. For details see the main text.
Model inversion using DCM

The parameters in both the neural state model and the observation model were inferred from VSDI data using a standard variational Bayesian scheme under Laplace approximation (Friston et al., 2007; Friston et al., 2003), implemented in the SPM12 toolbox (https://www.fil.ion.ucl.ac.uk/spm/, spm_nlsi_GN.m). The details for DCM inversion steps can be found elsewhere (Friston et al., 2007; Friston et al., 2003) and briefly in Appendix 1.

After estimating DCM for each individual, the model evidence was evaluated at the group-level using parametric empirical Bayesian modeling (PEB) for DCM (Friston et al., 2015; Friston et al., 2016). The PEB for DCM is designed to account for random (between session and subject) effects on parameters that are estimated during the first (within session or within subject) level of DCM. More specifically, the observation of the $i$-th subject, $y_i$, in the first level (within-subject), is described by a function $\Gamma(\theta)$ of its parameters $\theta$ and estimation error $\epsilon_i^{(1)}$. At the second level (between-subject), individual DCM parameters are modelled with a group design matrix $X$, between-subject effects $\beta$ and random effect $\epsilon^{(2)}$,

$$
y_i = \Gamma(\theta_i) + \epsilon_i^{(1)}, \quad (\text{Eq. 9})$$

$$\theta = X\beta + \epsilon^{(2)}. \quad (\text{Eq. 10})$$

The group-level approximation for parameters or connectivity $\beta$ can be obtained by using a design matrix of the second level $X=[1 1 1 1 \ldots 1]^T$ (N subjects $\times$ 1). We then have a group-level (approximated) model evidence (or free energy) for the observed data in a group, which was used as a model fit index for model comparison and model selection.

We used a free energy (an approximate log model evidence, See Appendix 1 for its definition) metric to evaluate model adequacy. We also used the percent explained variance (EV) as an index for the model fit of the experimental data.

$$EV = 100 \cdot \frac{\Sigma_{r,t}(y_i(t))^2}{\Sigma_{r,t}(y_i(t))^2 + \Sigma_{r,t}(y_{\text{model},r}(t)-y_{\text{model},r}(t))^2}, \quad (\text{Eq. 11})$$

in which $y_i(t)$ and $y_{\text{model},r}(t)$ respectively represent observation and estimated signal of a region $r$ at time $t$.

Model construction using a sequential Bayesian hyperparameter optimization
Experimental reference values for neural properties in the model would confine search space within a biologically plausible range. However, some neural properties do not have reference values nor have \textit{a priori} knowledge for their ranges. For this case, we introduced a method to approximate reference values by using a Bayesian optimization technique, which searches appropriate sampling points of hyperparameters based on a model of Gaussian process.

The Bayesian optimization of hyperparameters, which has been highlighted in the field of machine learning, tunes hyperparameters that determine a model structure (or configuration of model structure). In this scheme, we search a global maximizer of an objective function $F(\beta)$ (here, free energy function) with a hyperparameter vector $\beta$ ($\beta \in M$), within a model design space of interest $M$. Note the hyperparameter vector $\beta$, based on which parameters are estimated in DCM, are used as a parameter of the objective function $F$. To find optimal hyperparameters efficiently with the minimum number of steps (i.e., evaluations of the objective function, a computationally high-cost function), Bayesian optimization is composed of two steps - approximating the objective function (using a surrogate model) and an acquisition function to decide where to sample (evaluate). The surrogate model incorporates prior belief about the objective function and updates the prior with samples evaluated from the function to derive a posterior and thus leads to a better approximation for the function. Based on the posterior of the objective function, the acquisition function determines the next sample $\beta$ for evaluation that is expected to improve the approximation best over the currently accumulated evaluations. In the present study, we adopted a Gaussian process model for the surrogate model (Snoek et al., 2012) and an expected-improvement-per-second-plus function for the acquisition function (Bull, 2011; Gelbart et al., 2014). This approach is highly advantageous for cases with a high computational cost for model evaluation, relatively small dimension of hyperparameters, and unavailable hyperparameter gradients.

The optimal (unknown) reference values obtained after Bayesian hyperparameter optimization (and known reference values) are used as constants in the DCM model. As explained in Supporting Information S1, DCM model inversion is also a Bayesian parameter optimization process, using the variational expectation-maximization (EM) (Friston et al., 2007). Variational EM-based DCM estimates not only connectivity model parameters but also internal hyperparameters for controlling error and parameter precisions (denoted as $\lambda$, in the right upper panel of Figure 3). The hyperparameters $\lambda$ inside DCM differ from the hyperparameters designated to indicate unknown reference values, which are to be estimated before DCM model inversion. To differentiate it from the Bayesian parameter optimization technique used in DCM, we named the Bayesian optimization for exploring unknown
reference values as “Bayesian hyperparameter optimization” and utilized the MATLAB
function ‘bayesopt’ (Mathworks, co. USA) with a default acquisition function (‘expected-improvement-per-second-plus’). As a cost function, we used negative free energy
(approximate model evidence) of VSDI-DCM to be minimized by parameter updating.
Considering the high computational load in each DCM inversion using variational EM, this
approach is highly efficient in reducing the number of DCM inversion steps by utilizing
appropriate sampling of hyperparameters in a model driven approach (i.e., Gaussian process)
during the optimization.

In the current model for the hippocampus, 23 reference values are to be determined using
Bayesian hyperparameter optimization (see Tables 1 and 2). Since it is impractical to search
for an optimal value in a high dimensional space, spanned by all reference values at a time,
we propose a sequential optimization method of searching (optimizing) for a subset of the
reference values, followed by searching the other subset with the previously estimated sets
fixed (Figure 4). Subsets for the reference values were maximal PSPs of the excitatory and
inhibitory neural populations \( H_{exc} \) and \( H_{inh} \), modulation effects for external input signal \( C_0 \),
VSDI amplitude scaling factors for a region \( r \) (\( \alpha_r \), \( r=1,2,3 \)), contribution ratios \( \rho_i \) for each
neural population \( i \) (\( i=1,2,3 \) for CA1 and CA3 or 4 for the hilus) at region \( r \) and prior means
of intra and inter-regional connectivity \( A_{intra} \) and \( A_{inter} \) (Figure 4A).

\( V_{offset} \) for the reference PSP (in Eq. 6) and the slope \( R \) of the sigmoid function were also
estimated using the Bayesian hyperparameter optimization approach. The order of the subsets
(as well as categorization for subsets) was determined empirically; we first chose parameters
\( H_0 \) and \( C_0 \) during the first stage since those values changed (generated) signals dramatically.
After the sequential optimization, we finally performed Bayesian optimization for all
reference values \( (H_{exc}, H_{inh}, C_0, \alpha_r, \rho_i, A_{intra}, A_{inter}, V_{offset}, R) \) within the hyper-space search
range around the pseudo-optimal set of reference values in the sequential optimization.

We applied Bayesian hyperparameter optimization to the group averaged VSDI signal
after normalizing the mean amplitude of each slice (in the wild-type mice) to be in the level
of the maximal mean amplitude across slices. Scales of VSDI signals differ not only across
slices but also across groups. The scales of VSDI signals of the wild-type mice were
approximately two or three times larger than those of the mutant mice. Based on the reference
values found in the wild-type data, we sequentially estimated regional scaling factors for the
mutant group by estimating 1) the global scaling factor \( a_0 \) and 2) regional scaling factors
specific to each region \( r \) (\( \alpha_r \), \( r=1,2,3 \)) (step 6 in Figure 4A).
In summary, by introducing a sequential Bayesian hyperparameter optimization scheme to DCM, we found optimal reference values in high dimensional space, which reduces trial and errors in finding those values to construct a VSDI-DCM for the hippocampus.

**Figure 4.** A procedure for sequential Bayesian hyperparameter optimization. (A) In the Bayesian hyperparameter optimization, a total of 23 unknown reference values (unavailable from previous literatures) were sequentially optimized. (B) In each step, we searched 120 data points in the hyperparameter space, and compared the highest free energy to that of the previous step. Estimated parameters of the final step were used in the first-level (individual slice) model inversion using DCM, which optimizes the effective connectivity ($A_{ij}$) from neural population $j$ to $i$, the strength of input external stimuli ($C_i$) and the modulation parameter for maximal PSP ($H_i$) of the neural population $i$. (C) The cost function (the negative free energy) in step 1b is shown as an example of Bayesian hyperparameter optimization.

**Model reduction: decision for fixed and unfixed usage of reference values in the model**

In the VSDI-DCM, the reference values (adopted from previous literatures or estimated from Bayesian hyperparameter optimization) can be used as either fixed constants or unfixed variables to be estimated in the DCM. For the unfixed usage, the reference value is combined with a modulation parameter $\theta$ that controls an effect that has deviated from the reference value, by using either the exponential transformation $C_i = C_0 \exp(\theta_i)$ with $\theta_i \sim N(0, 1/32)$ or the absolute transformation $A_{ij} = \text{abs}(\theta_{ij})$ with $\theta_{ij} \sim N(A_{\text{intra/inter}}, 1/32)$, as listed in Table 2.
We now propose a method to determine whether to fix a reference value or to relax the effect as a variable using a modulation parameter. For example, we determined neural populations that do not necessarily need “free” maximal PSP $H_i$ to account for individual variability. For each slice of the wild-type, we estimated a full DCM model that comprises free (with modulation parameters $\theta_i H_i$) at all 10 neural populations, followed by Bayesian model reduction (BMR) (Friston and Penny, 2011). We generated $2^{10}$ combinations of reduced models nested in the full model by shrinking $H_i$ (setting both expectation and covariance to zero, e.g., $H_i = H_{exc} \exp(0)$ for an excitatory population or $H_i = H_{inh} \exp(0)$ for an inhibitory population). We then applied PEB for all reduced models to estimate group-level (approximated) model evidences (free energy) (Friston et al., 2016). If a population has a free $H_i$ in the optimal model with the highest group-level free energy, it indicates that $H_i$ may well be estimated to explain individual variance in the signal of neural population $i$ at the cost of increased complexity (Figures 3D and 4A).

Model reduction: decision for inclusion of potential connections in the model

In the hippocampus, some connections are not observed yet but may potentially exist (i.e., morphologically inferred connections) (Wheeler et al., 2015). To determine which combination of potential connections could be included in the model to explain the hippocampal responses better, we applied group-level BMR.

First, for each slice of wild-type mice, we estimated a full DCM model that comprises all known and potential connections. Using BMR, we generated reduced models nested in the full model by shrinking (thus removing) some of potential connections (setting both expectation and covariance to 0) (Friston and Penny, 2011). To find the best combination of connections to remove, we conducted a greedy search method at the group-level, by extending model reduction using greedy search at the individual-level DCM (Friston and Penny, 2011) to the group-level.

Since the full model of the present study includes 24 potential connections, it is not trivial to evaluate all $2^{24}$ reduced models. Therefore, by shrinking one-by-one the potential connections from the full model, we generated 24 reduced models. We then applied group-average PEB (with a vector of all ones for individuals in the design matrix $X$ in Eq. 10) for the reduced models to estimate group-level (approximated) model evidences (free energy) (Friston et al., 2016). The group-level free energies of the reduced models are compared with that of the full model and are used to determine whether to include (higher group-level free energy in the reduced model) or exclude (lower free energy in the reduced model) the
connection in the subsequent reduction of potential connections.

By doing this, we first selected the best eight (out of 24) connections that increased group-level free energy maximally after removal of those connections. Such connections indicate redundant contributions to the observed signal. Then, we constructed $2^8$ reduced models (binary combinations of 8 connections) and evaluated the group-level free energies for the reduced models using PEB. The reduced model with the highest free-energy was selected and the corresponding shrunken connections were removed. These steps were performed iteratively until the reduced models no longer improve free energy. This is an extension of the greedy search in conventional BMR implemented in `spm_dcm_bmr_all.m` of SPM12, which evaluates free energy at the individual level.

So far, we have described a group-level model selection (reduction) procedure using PEB by reducing (or shrinking) modulation parameters $\theta$ about neural properties (e.g., $H_i$) and by greedy searching of potential connections. We re-estimated DCMs with the reduced model configuration for each slice.

Bayesian group comparison of effective connectivity between the wild-type and mutant mice

For the mutant mice, after optimizing scale factors $\alpha$ for the average signal of the mutant mice, DCM inversion for individual slice was applied to a mutant-specific DCM model.

We then evaluated group-level differences in the parameters (e.g., effective connectivity) between wild-type and mutant mice using PEB. For this, we assigned a design matrix $X = [X_1 \ X_2]$ in which $X_1 = [1 \ 1 \ldots \ 1]^T$ ($N$ subjects of wild-type and mutant $\times$ 1) and $X_2 = [1 \ 1 \ldots \ 1 \ -1 \ -1 \ldots \ -1]^T$ (subjects of wild-type $\times$ 1 and mutant $\times$ -1) in Eq. 10. The parameter for the first column indicates the group common effect while the parameter for the second column indicates the group difference effect.

Perturbation analysis to evaluate causal effects of connections to mutant responses

Mutant mice had low VSDI signals and did not show the cumulative response patterns for consecutive stimuli that were observed in wild-type mice. Among many potential parameters that may explain in altered signals in mutant mice, we attributed altered response patterns (non-accumulated waveforms) of mutant mice to their altered effective connectivity. To focus on between-group differences in the accumulative waveforms (which we attribute to the effective connectivity differences), rather than signal amplitudes, we separated the effects of
scale factors $\alpha$ from the connectivity factors by estimating scale factors for each group.

As mentioned before, not all the connections that show group differences may be critical to induce altered responses in mutant mice. To test causality of the effective connectivity for altered responses in the mutant mice, we adopted a perturbation method proposed in Kang et al. (2017). We artificially perturbed the wild-type system by replacing some or all of the wild-type connectivity (showing significant group differences) with those of mutant mice without changing other parameters of the wild-type mice (except for scales), in order to generate mutant like VSDI signals.

First, we constructed a virtual model with parameters from the wild-type mice. Scaling factors of the observation model are derived from the mutant mice ($\alpha_r$, $r = 1, 2, 3$). Then, all possible combinations of connections (showing group difference) of wild-type mice were replaced with those of mutant mice. If the replacement generates similar signals to those of the mutant mice, we could argue that the replaced combination of connections is critical in explaining the underlying mechanism of the mutant mice. From ten group-differing connections detected using PEB analysis (the details will be explained in the Result section), we constructed 1024 ($2^{10}$) combinatory replacement sets of connections between the wild-type and mutant mice. We simulated VSDI signals for each replacement set, and compared the generated VSDI signals with those of the target mutant signals. We then evaluated at each region the ratio of EV for each replacement combination to EV for the wild-type without replacement, as follows.

$$\frac{EV_r}{EV_{0,r}} = 100 \times \frac{\sum \sigma(y_{\text{target},r}(t))^2}{\sum \sigma(y_{\text{target},r}(t))^2 + \sum \sigma(y_{\text{target},r}(t) - y_{\text{model},r}(t))^2}$$  (Eq. 12)

in which $y_{\text{target},r}(t)$, $y_{\text{wt},r}(t)$, and $y_{\text{model},r}(t)$ designate signals of a region $r$ at time $t$ generated with model parameters of the target mutant, wild-type, and replaced model parameters. To focus on the immediate responses for stimuli, we restricted the time range of evaluation from 0 to 500 ms.

Table 3 summarizes all the procedures for VSDI-DCM for the hippocampus.

| Table 3. Summary of all procedures for VSDI-DCM |
|-----------------------------------------------|
| Procedures* | Relevant information and results |
| Construction of a neural state model: | |
| Assign (known) reference values available from literatures and the Hippocamrome Database | Figure 3A, Table 1 |
| $V_{\text{rest},r}$, $V_{\text{th},r}$, $f_r$, $\tau$, $R$ | |


Bayesian hyperparameter optimization: Decision of optimal reference values for unknown neural properties in wild-type mice

$H_{exc}$, $H_{inh}$, $\rho$, $C_0$, $V_{offset}$, $R$, prior mean of $A_{intra}$ and $A_{inter}$, and $\alpha_i$ using averaged VSDI data of the scaled signals of wild-type mice

Unknown reference values are chosen.
Figure 3B and 4, Tables 1 and 2

Bayesian hyperparameter optimization: Decision of an optimal reference value specific to the mutant, $\alpha_r$, using the wild-type reference values except for $\alpha_r$ for the mutant.

Reference value specific to the mutant was chosen.
Figure 3C and 4, Tables 1 and 2

Estimation of DCM parameters for each individual:
Adopting hyper-parameters that were determined in the above steps, DCM parameters $[A_{ij}, C_i, H_i]$ for each slice were estimated.

DCM parameters of $[A_{ij}, C_i, H_i]$ for each slice, and their free energy.
Figure 7

Determination of maximal PSP $H_i$ as fixed or free in DCM using group-level model comparison

Group-level model comparison (PEB) of models with maximum PSP $H_i$ as a free parameter and reduced models with $H_i$ as a constant in the DCM. All possible 1024 combinations (on/off of maximal firing rate of 10 nodes) were evaluated.

Two $H_i$ were chosen to be free.
Figure 3D and Figure 8

Selection of the best connection model using a greedy search algorithm for the best group model evidence.
Selection of the best model after greedy search for all reduced models, utilizing PEB.

A full model with all the potential connectivity was chosen.
Figure 3E and Figure 8

Group comparison of effective connectivity using PEB

Figure 3F and Figure 9

Perturbation analysis for causality of the effective connectivity

Figure 3G, Figure 10

* Index of $i$ and $r$ represent id of neural population and region, respectively.

RESULTS

Improved model estimation by Bayesian hyperparameter optimization

Sequential Bayesian hyperparameter optimization increased the model evidence for the average VSDI signal of the wild-type mice (Figure 5). The EV increased from 91.4 % to 98.6 % by using Bayesian hyperparameter optimization. A simple simulation is also presented in Supporting information S1 to show the performance improvement by using Bayesian hyperparameter optimization, compared to a one-step DCM that estimates all reference values (hyperparameters) and parameters simultaneously within the conventional DCM. Bayesian hyperparameter optimization for reference values of intra- and inter- regional effective connectivity ($A_{intra/inter}$) found optimal values at 0.79 and 0.39, which were used as prior...
expectations (with prior covariance of 1/32) for the intra- and inter-regional effective connectivity $A_{ij}$ in the neural state model in Eq.5. (see Table 2).

The average VSDI signal of the mutant group was roughly four times weaker than that of the wild-type group. When we estimated scaling factors for all regions of the mutant mice as well as those of the wild-type mice, the scaling factors of the mutant mice were 0.3-0.4 times lower than those of the wild-type mice; $[\alpha_1, \alpha_2, \alpha_3]_{\text{wt}} = [1.30, 0.44, 0.18]$ and $[\alpha_1, \alpha_2, \alpha_3]_{\text{mut}} = [0.50, 0.15, 0.06]$ (Table 1).

Based on the parameters estimated from group average signals, we performed DCM inversion for individual slices. EVs of individual slices were 84.5 – 95.1 % (mutant), and 93.3 – 98.3 % (wild-type).

**Figure 5.** Bayesian hyperparameter optimization for reference values using the group average signals of the wild-type mice. The predicted signals before and after the Bayesian hyperparameter optimization are displayed in (A) and (B), respectively. Dotted lines and solid lines respectively indicate observed and predicted VSDI signals. Blue, red, yellow lines represent VSDI signal at the hilus, CA3 and CA1 regions, respectively.

**Parameterization of the membrane potential-to-firing rate transforming sigmoid function**

Using Bayesian hyperparameter optimization, we obtained $R = 0.21 \text{ mV}^{-1}$ and $V_{\text{offset}} = 7.1 \text{ mV}$ for the sigmoid function (membrane-potential to firing-rate). To gain some insight about these values, we evaluated the effects of a sigmoid function with regard to two parameters $R$ and $V_0$ by simulating neural responses with the previous fixed parameter set ($R = 0.67 \text{ mV}^{-1}$) (Moran et al., 2013) and the parameter set obtained from the Bayesian hyperparameter
optimization ($R = 0.21 \text{ mV}^{-1}$, $V_{\text{offset}} = 7.1 \text{ mV}$). We constructed a simple computational model which consists of pyramidal (E31) and basket (I31) neuronal populations at CA3. We set one connectivity from E31 to I31, $A_{31,31} = 0.3$. Four electric pulse stimuli with 100 ms interval were applied on the pyramidal neural population with the parameters of the sigmoid function: $R = 0.67 \text{ mV}^{-1}$ and $V_0 = V_{\text{th}} (-50.0 \text{ mV})$, and with the parameters from the Bayesian hyperparameter optimization: $R = 0.21 \text{ mV}^{-1}$ and $V_0 = V_{\text{th}} + V_{\text{offset}} = -42.9 \text{ mV}$. To evaluate the effect of each parameter for different input strengths (controlled by changing modulation parameter $\theta$ with -1.6, -0.8, 0.0, 0.4, and 0.8), we performed simulations of four models: 1) $R = 0.21 \text{ mV}^{-1}$, $V_0 = -42.9 \text{ mV}$; 2) $R = 0.67 \text{ mV}^{-1}$, $V_0 = -42.9 \text{ mV}$; 3) $R = 0.21 \text{ mV}^{-1}$, $V_0 = V_{\text{th}} = -50.0 \text{ mV}$; and 4) $R = 0.67 \text{ mV}^{-1}$, $V_0 = -50.0 \text{ mV}$.

The sigmoid functions with the estimated slope ($R = 0.21 \text{ mV}^{-1}$) shows a wider range of monotonic increase in the range of -60 mV to -20 mV compared to those with $R = 0.67 \text{ mV}^{-1}$ (Figure 6B). Increased $V_0$ parameter ($V_0 = -42.9 \text{ mV}$) shifted the monotonic increase range to a higher voltage range. The model with $R = 0.21 \text{ mV}^{-1}$, $V_0 = -42.9 \text{ mV}$ showed a wider and increased dynamic range of responses at I31 for different strengths of inputs (Figure 6C).

A further simulation result presented in Supporting information S2 suggests the need for hyperparameter optimization of $R$ and $V_{\text{offset}}$ when applied to real data set; this is explained in the following section.

**Figure 6.** Effects of four firing-rate transfer sigmoid functions. A simple system, which simulates a pyramidal (E31) and a basket (I31) neuronal population of the CA1 region, is shown in (A). Sigmoid functions of E31 evaluated are shown in (B). The membrane potentials ($V_m$) of E31 and I31 are shown in (C). The inputs to E31 were weighted by modulation values, $C_0 \exp(\theta)$. For simulation, we used five different modulation parameters $\theta$ of -1.6, -0.8, 0.0, 0.4, and 0.8 to induce different amplitudes of input effects. Four models with different sigmoid functions are $R = 0.21 \text{ mV}^{-1}$ and $V_0 = -42.9 \text{ mV}$ (solid red); $R = 0.67 \text{ mV}^{-1}$ and $V_0 = -42.9 \text{ mV}$ (solid blue); $R = 0.21 \text{ mV}^{-1}$ and $V_0 = V_{\text{th}} = -50.0 \text{ mV}$ (dotted red); $R = 0.67 \text{ mV}^{-1}$ and $V_0 = V_{\text{th}} = -50.0 \text{ mV}$ (dotted blue). The purple rectangle corresponds to the
range between an action potential threshold (-62.8 mV) and a maximal membrane potential (-27.3 mV) at $\theta_i = 0$ in (C).

**Bayesian model reduction**

Figure 7A and Figure 7B present examples of predicted signals at a wild-type slice and a mutant slice, respectively. Here, the effective connectivity ($A_{ij}$), input strength ($C_i$), and maximal PSP ($H_i$) were allowed to be different across individual slices. Free parameterization of $H_i$ was implemented by utilizing modulation parameter $\theta_i$ that controls an effect deviating from the reference value, i.e., $H_i = H_{exc} \exp(\theta_i)$ for excitatory neural populations or $H_i = H_{inh} \exp(\theta_i)$ for inhibitory neural populations, with $\theta_i \sim N(0,1/32)$ (see Table 2). Both $H_{exc}$ and $H_{inh}$ were obtained via Bayesian hyperparameter optimization. The modulation parameter $\theta_i$ of $H_i$ for each neural population did not significantly differ from zero (95% credibility interval belonging to zero), i.e., $H_i \times \exp(0)$, in these examples of both the wild-type and the mutant mice.

**Figure 7.** Representative examples of individual DCM analysis of a wild-type (A) and a mutant mouse (B). The predicted VSDI with a model of free parameters [$A_{ij}$, $C_i$, $H_i$] in the DCM estimation is presented for three hippocampus regions. Blue, red, yellow lines in the left column represent VSDI signal at the hilus, CA3 and CA1 regions, respectively. Dotted lines and solid lines show observed and predicted VSDI signals. Expectation of parameters are presented in gray bars with a credible interval of 95% in red bars.

Next, we conducted group-level model reduction to determine whether to assign $H_i$ for each population as a fixed constant or as a free parameter (variable across individuals). The concept and results of the group-level model reduction with multiple combinations of on-off $H_i$ across neural populations are presented in Figure 8A and 8B. Group-level model reduction
found that the best model among 1024 ($2^{10}$) models of all possible on-off $H_i$ combinations was the reduced model with all fixed $H_i$ but one free $H_i$ ($H_{E12}$) (Figure 8A and Figure 8B). Therefore, a model with one free parameterized $H_i$ ($H_{E12}$) was used in the subsequent analysis of individuals.

Figure 8. Bayesian model reduction for maximal PSP $H_i$ (A and B) and model reduction for potential connections using a greedy search algorithm (C and D). All possible combinations of fixed/unfixed usage of the maximal PSP $H_i$ were modeled (A) and their group-level free energies derived from PEB were evaluated (B). Free energies of all possible models (left) and the best 11 models (right) for the maximal PSP $H_i$ are displayed (B). Combinations of potential connections were shrunken by setting the prior mean and covariance of each connection to zero (C). The free energy of 256 models and the best 11 models after a greedy search to remove redundant connections are displayed in (D).

To evaluate potential connections (connections presumed without solid references but morphologically inferred), we further conducted model optimization by evaluating the gain (in model evidence) of including potential connections using a greedy search algorithm. Figure 8C and Figure 8D display the concept and the result of the greedy search algorithm that explored the best combination of potential connections for the hippocampus model, respectively. Although some reduced models among the 256 final models after the greedy search showed relatively higher group-level free energies than others, the full model showed the highest free energy (Figure 8D). Therefore, we selected the full model, which includes all potential connections in the model for the subsequent analysis. Although the using greedy search algorithm did not result in reducing model parameters in this particular example, it did provide a rationale for the usage of all possible connections in the hippocampus model. This result also supports the prediction raised by Peter’s rule (morphologically predicted
connectivity) (Rees et al., 2017).

PEB analysis between wild-type and mutant mice

Figure 9 displays PEB results for group common and differences in the model parameters. In the group common connectivity (Figure 9A and 9C), the strength of the intra-regional connectivity was generally higher than that of the inter-regional connectivity.

PEB identified ten significantly different effective connectivity between groups in a Bayesian criterion (with a 95% credible interval) (Figure 9B and 9D). The three group-differing inter-regional connections were from HIPROM of the hilus to the pyramidal neural population of the CA3 region (I12 → E21); from the pyramidal neural population of the CA3 region to the pyramidal neural population of the CA1 region (E21 → E31); and from the pyramidal neural population of the CA3 region to the recurrent O-LM population (E21 → I32). All inter-regional connections along the pathway from the hilus to the CA1 region were mediated by the pyramidal neural population of the CA3 region (E21). Three group-differing intra-regional connections in the CA1 region were associated with the O-LM population (I32): self-recurrent inhibition at the O-LM (I32), reciprocal connections between the O-LM and the pyramidal neural populations (I32 → E31 and E31 → I32). The self-recurrent inhibition at the O-LM (I32) in the CA1 region, self-recurrent excitation at the pyramidal population (E21) in the CA3 region, and inter-regional connection from the pyramidal population (E21) in the CA3 region to the pyramidal population (E31) in the CA1 region showed decreased effective connectivity in the mutant mice as compared to the wild-type mice. Except for these three connections, the other seven connections showed increased connectivity in the mutant mice compared to the wild-type mice.

Except for two excitatory-excitatory connections from E21 (E21 → E21 and E21 → E31), the connectivity showing group difference is associated with inhibitory effects, either with modulation from inhibitory neural populations or projection to inhibitory neural populations. Based on this result, the mutant mice seem to have hyper-inhibition among neural populations in processing the consecutive stimuli, which may result in reduced signals and no accumulative VSDI signals.
Figure 9. PEB results for group comparison. Group common and group difference in the effective connectivity are displayed with bar plots (A and B) and illustrations (C and D). Effect sizes are presented in gray bars with a credible interval of 95% (red bars) in A and B. Excitatory and inhibitory connections are colored with red and blue. Solid and dotted lines in (D) present stronger effective connectivity in the wild-type and mutant mice, respectively. The neural populations are; E11 (granule), E12 (mossy), I11 (DG Basket), I12 (HIPROM), E21 (CA3c Pyramidal), I21 (CA3 Basket), I22 (CA3 O-LM), E31 (CA1 Pyramidal), I31 (CA1 Basket), and I32 (CA1 recurrent O-LM).

Causal effects of effective connectivity: a perturbation analysis

For the causality test, we performed a perturbation analysis for effective connectivity in generating mutant-like signals (Figure 10). Causal effects of all possible combinations (Figure 10B) for the ten group-differing connections (Figure 10A) are presented in Figure 10D. As shown in Figure 10C, replacement of all ten connections with those of the mutant mice (and any combinations of the ten connections) was not sufficient to reproduce mutant-like responses at the hilus (Figure 10C, 10D, and 10E). This implies the necessity for additional connections other than just the group-differing connections to induce abnormal response at the hilus of the mutant mice. On the contrary, mutant-like abnormal responses in the CA3 region were achieved even by replacing a single connectivity from I12 to E21 (combinations of replacement 1 to 256 in Figure 10D); any combinations of replacement including the I12 to E21 connection increased the ratio of EV/EV₀ (to explain mutant-like signals) in the CA3
In the CA1 region, however, the effects of replacement combinations were complex (Figure 10D). The ratio of $EV/EV_0$ in the CA1 region increased with the number of replaced connections (Figure 10G). Except for the signal at the hilus, mutant-like responses could be obtained by replacing only four connections ($I_{12} \rightarrow E_{21}, I_{32} \rightarrow E_{31}, E_{21} \rightarrow I_{32}, I_{32} \rightarrow I_{32}$), not necessarily requiring replacement of ten connections (Figure 10C).

Figure 10. Evaluation of the causal role of each effective connectivity by replacement of wild-type connections with mutant connections. (A-B) Ten effective connectivity that showed group differences in PEB analysis (A) and all possible combinations of those connections to be replaced (B) are presented. (C) Simulated VSDI signals at the hilus, CA3 and CA1 regions are presented in blue, red, and yellow. The dotted lines and black solid lines represent target mutant signals and wild-type signals, respectively. Colored solid lines in the middle and bottom panels in (C) show signals that are generated with a model with four replaced connections ($I_{12} \rightarrow E_{21}, I_{32} \rightarrow E_{31}, E_{21} \rightarrow I_{32}, I_{32} \rightarrow I_{32}$) and a model with ten replaced connections. (D-G) The ratio of $EV/EV_0$ at the three regions for all possible combinations of replacements are displayed in (D). The ratio of $EV/EV_0$ according to numbers of replaced connections at the hilus, CA3 and CA1 regions are shown in (E-G). The effective connectivity $I_{12} \rightarrow E_{21}$ (asterisked in A) significantly increased $EV/EV_0$ of CA3.
DISCUSSION

Bayesian hyperparameter optimization

The success of computational modeling lies in striking a right balance between the simplicity and biological plausibility of the model. In construction of a biologically plausible model, the reference values for neurobiological properties play critical roles not only in confining model solutions within the biological ranges but also in simplifying the model to be estimated. However, not all neurobiological reference values are available. Furthermore, even the existing reference neurobiological properties may not necessarily be true for different animals in different context without adjustment. Thus, we proposed a two-step Bayesian optimization method by performing Bayesian hyperparameter optimization for model reference values, followed by optimization of model parameters (e.g., effective connectivity) in DCM. The Bayesian hyperparameter optimization we propose searches optimal reference values to maximize group-level model evidence as an efficient way of reducing the number of DCM inversions in the optimization. Although the computational cost for EM optimization during DCM inversion may not be high for a small network, the optimization of an unknown reference for DCM requires iterative evaluations with different samples of the reference. When multiple reference values have to be estimated, it is not trivial to conduct optimization in all multi-dimensional spaces (reference values) at a time. Thus, we proposed a sequential optimization approach of searching (optimizing) a subset of reference values, followed by searching the other subset, in order. Then, we confine the search space around optimal values for each parameter over which Bayesian optimization for all the reference values had been conducted. This sequential Bayesian hyperparameter optimization approach makes it practical to utilize Bayesian optimization to a problem in very high-dimensional space. We found improved model fit results by using Bayesian hyperparameter optimization (in both the simulation and experimental results, see also Supporting information S1).

The order of reference values to be estimated in the sequential optimization inevitably bias the optimization unless the reference values are independent. In this study, we performed the sequential Bayesian hyperparameter optimization by using an empirically determined order, which was determined by ordering the amount of free energy changes after changing each hyperparameter by trial and errors. Although we conducted Bayesian hyperparameter optimization with all reference values within a restricted range of previously determined optimal values to reduce such order-dependent effects at the final stage, we cannot rule out the possibility of a better solution that may be found by considering all reference values together from the beginning at high computational cost. Determining the order of reference
values for optimization will be a challenging problem and should be addressed in future studies.

Adjusted sigmoid firing-rate transfer function in the neural population dynamics

For the neural state model of ensemble activity, we used a convolution-based model that comprises inter-modal connectivity with a sigmoid transform function that converts the pre-synaptic membrane potential to the firing rate for each population and a synaptic kernel mapping from the firing rate to the post-synaptic membrane potential (Jansen and Rit, 1995; Marreiros et al., 2008; Moran et al., 2013). Since the convolution-based neural state model focuses on the probabilistic information transfer between neural populations, the ensemble activity of neurons is generally described by the mean property of neurons in the population. Thus, the firing-rate transfer sigmoid function ($\sigma(v)$ in Eq. 1) converts the presynaptic mean membrane potential to the mean firing rate of a local population (Liley et al., 2009; Marreiros et al., 2008; Marreiros et al., 2010; Spiegler et al., 2010). The mean postsynaptic potential is calculated in terms of convolution of a synaptic kernel with the mean firing rate of a neural population (called “neural mass”) (Freeman, 1975; Jack et al., 1975; Jansen and Rit, 1995; Wilson and Cowan, 1972). In practice, it is not easy to find references for the parameters of the firing-rate transfer function for each neural population, such as the PSP at the half-maximal firing rate $V_0$, maximal firing rate ($f_{max}$), and slope ($R$).

In the initial Jansen-Rit model (Jansen and Rit, 1995), the authors used 6 mV for $V_0$ and $R = 0.56 \text{ mV}^{-1}$ (adopted from Freeman, 1987) for EEG modeling (based on the local field potential with a distant reference), which may not be applicable to a membrane-potential-based model of the hippocampus (e.g., the membrane potentials of general neural cells around -60 ~ -80 mV may not achieve the level of 6 mV).

The firing-rate transfer function in a neural population can be described as a combination of population dynamics (proportion of firing neurons at the membrane potential) and a firing rate function of the membrane potential of each neuron. Due to the lack of sufficient references for the latter (indeed, derivation of the PSP at the half-maximal firing rate $V_0$ in a neuron is not trivial in practical experiments, as noted in Jansen and Rit (1995)), we estimated the parameters for the firing-rate transfer function by introducing Bayesian hyperparameter optimization. Since a PSP increases the firing rate of a neuron, the populational transfer of the average membrane potential to the mean firing rate is considered to operate approximately within a range of the average action potential threshold and a certain voltage above that threshold. Thus, we optimized the population reference PSP $V_0$ (inducing the 50% maximal
firing rate) at a higher voltage than the known action potential threshold. We also expected to find a wider range of monotonic increase from the action potential threshold to the upper bound of the membrane potential. As shown in Figure 6, the estimated parameter $R = 0.21 \text{ mV}^{-1}$ of the sigmoid function induces wider ranges of monotonic increase than $R = 0.67 \text{ mV}^{-1}$ used in Moran et al. (2013) and allows firing rates to depend more on the strength of input signals rather than binarized responses. The model with the optimized parameter set ($R = 0.21 \text{ mV}^{-1}$ and $V_0 = -42.9 \text{ mV}$) has a wider dynamic range of the sigmoid function that responds to a monotonic increase for different strengths of inputs as shown in Figure 6C. To reduce model parameters, we simplified the model by using the same sigmoid function with an offset from the action potential threshold (differing across neural populations) and a slope independent of brain regions. Although this approach works well as shown in Supporting information S2, these simplifications could be improved in further studies.

**Model simplification by a group-level model reduction**

After implementing Bayesian hyperparameter optimization for reference values, the decision for *fixed* and *unfixed* usage of each reference value in DCM depends on the hypothesis or question of interest. This is the question of whether or not to allow individual differences in the maximal PSP $H$ (as an example of a reference value) in explaining individual VSDI data. In this study, we proposed a method to determine whether to assign a modulation parameter for the reference value $H$ to the effect of $H$ as a variable to be estimated (*unfixed*) or simply use $H$ as a *fixed* constant in DCM by comparing second-level model evidence (approximated free energy) obtained from PEB analysis (Table 1). This approach can be a useful tool when no clear hypothesis for the effects of inter-individual variability of each parameter is available.

For the neural circuit of the hippocampus, we constructed a full model with all known and potential connections based on a hippocampus database. In the hippocampus, the potential connections are assumed to exist based on Peter’s rule (morphologically predicted connectivity) (Rees et al., 2017), but the functionality of these connections has not been fully evaluated. To test the need for a potential connection in the processing of stimuli, we performed a Bayesian model reduction of combinations of connections (Friston et al., 2016), in a greedy search manner. The greedy search-based model reduction suggests that the full model is the best model for explaining the observed VSDI signals. This is a good example that Bayesian model comparison may contribute to testing plausible circuitry for which no anatomical evidence is yet available.

In a technical perspective, the novelty of the current procedure (testing the need for potential connections) is the introduction of a greedy search method for group-level model evidence, which extends the greedy search-based model reduction at an individual-level DCM
(Friston and Penny, 2011). In particular, evaluation of model evidence for each connection on a group-level using PEB is advantageous when modeling noisy signals as well as noisy individuals (variation from group mean). Since the second-level model regularizes estimation of parameters in the context of both the first and second-level variations, we speculate that group-level evaluation is more robust to estimation errors than evaluation at an individual level. Note that the current method is similar to the hierarchical model comparison method implemented in `spm_dcm_peb_bmc.m`, except for the order of implementation. The proposed method conducts model reduction (during the greedy search) in the first (individual) level and test the free energy for the group-average of the reduced model using PEB at the second level. On the other hand, `spm_dcm_peb_bmc.m` constructs a group-averaged full model using PEB and conducts greedy search for reduced models nested in the full model in the second level. The cons and pros of each method remain to be evaluated in future studies.

### Computational models for the hippocampus and DCM

As described in the introduction, the hippocampus is a compelling target for the computational modelling, particularly on the microscopic circuit level. Bezaire et al. (2016) performed perturbation analysis using a realistic full-scale model of the rodent CA1, which includes one excitatory neuron type (pyramidal cells) and eight inhibitory (interneuron) types, and suggested that parvalbumin-expressing interneurons, neurogliaform cells, and interneuronal diversity are important for generating intrinsic theta rhythm (Bezaire et al., 2016). Sanjay et al. (2015) conducted a computational model study of the temporal lobe epilepsy. By modeling the CA3 region using pyramidal cells, basket cells, and O-LM cells, they investigated how loss of connectivity between the O-LM and pyramidal cells influences the behavior of other types of neurons, how changes in external inputs received by pyramidal neurons affect their activity, and how changes at synapses between the neurons affect the overall network activity in the CA3 subfield (Sanjay et al., 2015). The authors concluded that impaired dendritic inhibition and changed synaptic plasticity can induce epileptic activity.

Most of these computational modelling studies have focused on modelling intra-regional connectivity at a specific region of the hippocampus. Furthermore, these complex computational models were not designed to fit individual data and allow for inter-individual variations. By introducing a computational model for the hippocampus on a mesoscopic level using the DCM framework, we were able to explore stimulus-dependent intra- and inter-regional effective connectivity in the hippocampus of a group of wild-type mice and another group of mutant mice. Although the current computational model of the hippocampus contains relatively small numbers of neural populations and model parameters (compared to microscopic models), this approach is advantageous in exploring group-specific neural mechanisms using observed data.
The neural state model for the hippocampus we introduce is based on a well-known convolution-based model, which has widely been used for modeling effective connectivity among cortical columns (Bastos et al., 2012; Jansen and Rit, 1995; Moran et al., 2013). In contrast to computational modelling of cortical columns, most previous computational models of the hippocampus have employed a single principal neuron (excitatory neuron) and multiple inhibitory neurons for each column and focused on the functional role of inhibitory neurons (Bezaire et al., 2016; Kullmann, 2011; Sanjay et al., 2015; Wendling et al., 2002). In line with those studies, we include one excitatory (or two for the hilus) and two inhibitory neural populations to construct a region in the hippocampus.

For the hippocampal inter-regional circuit, we included three neural populations of those dendrite that are located in the stratum oriens (and not in the stratum radiatum) as latent neural populations (I22 and I32 in Figure 2), although the dendrite locations of these hidden neural populations are not in the region of interests for VSDI measurement. This is because the functional role of various interneurons, including basket and O-LM neurons, has been widely investigated (Bezaire et al., 2016; Sanjay et al., 2015). To include these unobserved neural populations in the model, we take advantage of the Bayesian scheme of DCM to approximate latent variables under constraints of the model topology.

It should be noted that the time constant of the synaptic kernel, particularly for inhibitory neural populations, is one of the key parameters in the epileptic neural circuit. In the current model, the two inhibitory neural populations at the hilus and CA3 have different levels of time constants, i.e., faster (~10 ms, I11 and I21) and slower ones (~30 ms, I12 and I22), except at CA1 that has two different levels of membrane potentials and action potential thresholds (Table 1). In the model, the slower inhibitory neurons (I12 and I22) were configured to project to faster inhibitory neurons (I11 and I21) (Figure 2). To focus on connectivity alterations reflected in the accumulative waveforms and to reduce the number of parameters, we did not estimate the time constants that are relatively well documented in the public database (Table 1). However, the systematic evaluation of interplay between fast and slow inhibitory neural populations should be further researched particularly in the mutant for epileptic mice, exampled in Wendling et al. (2002).

Altered connectivity in the knock-out mouse (Arx\textsuperscript{+/};Dlx5/6\textsuperscript{CIG})

ARX, which is located on the X-chromosome, is an important transcription factor for the development of the forebrain, pancrease, and testes development (Gecz et al., 2006; Olivetti and Noebels, 2012). In the brain, it is expressed in the subpallial proliferative zones and in developing and adult interneurons. Many ARX mutations have been reported in brain diseases such as X-linked infantile spasms syndrome (ISSX) (Gecz et al., 2006; Olivetti and Noebels,
Marsh et al. (2009) developed a conditional Arx mutant mouse, where Arx is knocked-out in cells expressing Dlx5/6; Arx<sup>+/−</sup>;Dlx5/6<sup>Cre</sup> (female) and Arx<sup>−/−</sup>; Dlx5/6<sup>Cre</sup> (male). In both mouse models, developmental epilepsy was observed, resembling that of patients with Arx mutations. An immunohistochemistry study (Marsh et al., 2016) showed that young (postnatal day 14) Arx<sup>+/−</sup>;Dlx5/6<sup>Cre</sup> mice have different Arx expression in the DG (reduced Arx expression) and the CA1 region (increased Arx expression) of the hippocampus. Additionally, adult mutant mice showed reduced Arx expression in the CA3 region (not statistically signified). The study also revealed significantly increased neuropeptide Y expression and significantly decreased parvalbumin expression in adult mutant mice (Marsh et al., 2016).

In this study, we explored how regional alterations in the gene expression of the mutant mice affect altered interactions among neural populations, reflected in the altered response waveforms. Group comparison showed significant alterations in the ten connections including inter-regional interactions between the hilus and CA3, and between the CA3 and CA1 regions. Intra-regional connectivity alteration mainly occurred within the CA1 and hilus where significantly reduced Arx expression was found (Marsh et al., 2016). Although the Arx gene mutation is expected to change local neural compositions in the early stage of development, the current result implies that it may eventually affect inter-regional connectivity during the brain development.

In previous studies with rodents, abnormal inhibitory connections in the CA1 region, which is the initial region of the TA pathway, have been suggested to be capable of inducing seizures. For example, Ang et al. (2006) suggested that less inhibitions of the CA1 region in epileptic rats could induce epileptic signals in temporal lobe epilepsy. In their experiments, responses of the CA1 region evoked by stimulating the TA pathway were reduced in epileptic rats that were treated with scopolamine methyl nitrate, a temporal lobe epilepsy model. Johnson et al. (2014) also reported that mice with lateral fluid percussion injury, which also increases the risk of seizures, show hyperpolarization in the stratum oriens of the CA1 region when evoked by stimulation. By changing the positions of stimuli, they investigated the effects of neurons and suggested that induced traumatic brain injury reduced CA1 output through abnormal synaptic connections that projected to the cannabinoid-sensitive interneurons. Although the origins of abnormal systems in previous studies may differ from those in the current study, these results are consistent with our results with regard to the alteration with inhibitory connectivity in the CA1 region. The inhibitory neural population I32 in the CA1 region of the mutant mice in the present study receives increased effective connectivity from the intra-regional excitatory neural population E31 and inter-regional effective connectivity neural population E21.

The alteration in the effective connectivity of the mutant mice was expressed in a cascaded manner; e.g., the bidirectional interaction of E31 and I32 in the CA1 was modulated
by E21 in the CA3, which was again affected by the inhibitory neural population I12 in the hilus. The details for these cascaded alterations and the causes for such alterations (either from primary alterations due to gene modification or secondary compensatory alterations) in the inter-regional connectivity, in association with intra-regional connectivity, should be further explored.

Statistical differences and their causality

PEB could identify group differences in effective connectivity. However, we showed that the statistical differences do not necessarily indicate a causal role of the significant connection in inducing abnormal responses in the mutant. To evaluate the causality of each connection, we took advantage of computational modeling in generating signals by artificially perturbing specific connections. From the perturbation analysis, we found three characteristics in the connectivity for mutant mice. First, the wild-type signal at the hilus was not altered by changing any possible combinations of the ten connections. This indicates that the signals of the hilus in the mutant mice may possibly be induced by a more complex circuit including additional connections that were not detected at the group comparison analysis. Second, perturbation analysis showed that only a single connection (from I12 to E21) was sufficient to induce mutant-like behavior in the CA3 region. Third, in contrast to the CA3 region, the altered response in the CA1 region may be attributable to the multiple combinations of the connections as the mutant-like behaviors was found according to the increased number of altered connectivity. These results suggest the importance of perturbation analysis in identifying causal relations of each connectivity or combinations of multiple connections in inducing mutant-like behaviors.

In summary, we proposed a Bayesian hyperparameter optimization scheme for determining hyperparameters used in DCM, and found increased performance in determining a mesoscopic computational model for the hippocampus by using our optimization scheme. We also proposed a VSDI-DCM scheme to investigate hippocampal activities at the neural population and analyzed the VSDI data of rodent hippocampal brain slices of wild-type and epileptic mutant mice. To achieve construction of a biologically feasible model, we extended the conventional DCM model by introducing data-based neural populations and hidden brain region in the circuits. As a result, VSDI-DCM successfully fits VSDI signals in both wild-type mice and epileptic Arx conditional knock-out mutant mice. Group comparison suggested altered effective connectivity mainly with inhibitory neural populations in the mutant mice. The perturbation confirmed the causal relations of the connectivity in inducing mutant-like neural responses. Although the current method was proposed to explain a special case of Arx
In Dlx5/6Cig knock-out mice, this model construction and optimization approach could be extended to diverse mesoscopic applications that require biologically sophisticated computational modeling using DCM.

Acknowledgments

This research was supported by Brain Research Program and the Korea Research Fellowship Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science and ICT (NRF-2017M3C7A1049051 and NRF-2017H1D3A1A01053094). The authors thank Prof. Karl Friston for the methodological guides and Minho Kim for his help on editing.

Appendix 1. DCM inversion

The variational expectation-maximization (EM) estimates model parameters $\theta$ for observed signals $y$ in the way of maximizing the log-evidence ($\ln p(y|m)$). The model log-evidence can be formulated in terms of KL-divergence between the posterior distribution $p(\theta|y,m)$ and its approximate distribution $q(\theta)$ and the free energy evaluated under $q(\theta)$,

$$\ln p(y|m) = KL\left(q(\theta)||p(\theta|y,m)\right) + F(q,y), \quad \text{(Eq. A1)}$$

$$F(q,y) = \int q(\theta) \ln \frac{p(y,\theta)}{q(\theta)} d\theta = KL\left(q(\theta)||p(y,\theta|m)\right). \quad \text{(Eq. A2)}$$

The free energy $F$ can be considered as an approximate lower-bound for log-evidence. The objective of optimization is to derive an approximate distribution $q(\theta)$ that maximizes $F$, thus results in $q(\theta) \approx p(\theta|y,m)$. Under the Laplace approximation, $q(\theta)$ are updated iteratively using an EM algorithm, which consists of expectation (E) and maximization (M) steps. In the expectation step, the conditional expectations and covariances of parameters $\theta$ are estimated by maximizing the free energy $F$, with given hyperparameters for error and parameter covariance components (denoted as $\lambda$ in Friston et al. (2007), see the upper right panel of Figure 2), using a Gauss-Newton algorithm. In the maximization step, the hyperparameters for error and parameter covariance components are updated in the same way to maximize free energy, based on the parameters estimated in the E-step (Friston et al., 2007).
REFERENCES

Ali, A.B., Thomson, A.M., 1998. Facilitating pyramid to horizontal oriens-alveus interneurone inputs: dual intracellular recordings in slices of rat hippocampus. The Journal of Physiology 507, 185-199.

Allen, K., Monyer, H., 2015. Interneuron control of hippocampal oscillations. Curr Opin Neurobiol 31, 81-87.

Ang, C.W., Carlson, G.C., Coulter, D.A., 2006. Massive and specific dysregulation of direct cortical input to the hippocampus in temporal lobe epilepsy. J Neurosci 26, 11850-11856.

Bastos, A.M., Usrey, W.M., Adams, R.A., Mangun, G.R., Fries, P., Friston, K.J., 2012. Canonical microcircuits for predictive coding. Neuron 76, 695-711.

Berger, T., Borgdorff, A., Crochet, S., Neubauer, F.B., Lefort, S., Fauvet, B., Ferezou, I., Carleton, A., Lüscher, H.-R., Petersen, C.C.H., 2007. Combined Voltage and Calcium Epifluorescence Imaging In Vitro and In Vivo Reveals Subthreshold and Suprathreshold Dynamics of Mouse Barrel Cortex. 97, 3751-3762.

Bezaire, M.J., Raikov, I., Burk, K., Vyas, D., Soltesz, I., 2016. Interneuronal mechanisms of hippocampal theta oscillations in a full-scale model of the rodent CA1 circuit. Elife 5.

Booker, S.A., Vida, I., 2018. Morphological diversity and connectivity of hippocampal interneurons. Cell Tissue Res 373, 619-641.

Bourgeois, E.B., Johnson, B.N., McCoy, A.J., Trippa, L., Cohen, A.S., Marsh, E.D., 2014. A toolbox for spatiotemporal analysis of voltage-sensitive dye imaging data in brain slices. PLoS One 9, e108686.

Braak, H., Braak, E., 1991. Neuropathological stageing of Alzheimer-related changes. Acta Neuropathol 82, 239-259.

Buckmaster, P.S., Strowbridge, B.W., Schwartzkroin, P.A., 1993. A comparison of rat hippocampal mossy cells and CA3c pyramidal cells. 70, 1281-1299.

Bull, A.D., 2011. Convergence rates of efficient global optimization algorithms. arXiv, 1101.3501.

Chemla, S., Chavane, F., 2010a. A biophysical cortical column model to study the multi-component origin of the VSDI signal. Neuroimage 53, 420-438.

Chemla, S., Chavane, F., 2010b. Voltage-sensitive dye imaging: Technique review and models. J Physiol Paris 104, 40-50.

Cutsuridis, V., Graham, B.P., Cobb, S., Vida, I., 2018. Hippocampal Microcircuits.

Eichenbaum, H., Cohen, Neal J., 2014. Can We Reconcile the Declarative Memory and Spatial Navigation Views on Hippocampal Function? Neuron 83, 764-770.
Fernandez, F.R., Broicher, T., Truong, A., White, J.A., 2011. Membrane Voltage Fluctuations Reduce Spike Frequency Adaptation and Preserve Output Gain in CA1 Pyramidal Neurons in a High-Conductance State. The Journal of Neuroscience 31, 3880.

Freeman, W.J., 1975. Mass Action in the Nervous System: Examination of the Neurophysiological Basis of Adaptive Behavior through the EEG. Academic Press, New York.

Freund, T.F., Buzsáki, G., 1996. Interneurons of the hippocampus. Hippocampus 6, 347-470.

Friston, K., Mattout, J., Trujillo-Barreto, N., Ashburner, J., Penny, W., 2007. Variational free energy and the Laplace approximation. Neuroimage 34, 220-234.

Friston, K., Penny, W., 2011. Post hoc Bayesian model selection. Neuroimage 56, 2089-2099.

Friston, K., Zeidman, P., Litvak, V., 2015. Empirical Bayes for DCM: A Group Inversion Scheme. Frontiers in Systems Neuroscience 9.

Friston, K.J., Harrison, L., Penny, W., 2003. Dynamic causal modelling. Neuroimage 19, 1273-1302.

Friston, K.J., Litvak, V., Oswal, A., Razi, A., Stephan, K.E., van Wijk, B.C.M., Ziegler, G., Zeidman, P., 2016. Bayesian model reduction and empirical Bayes for group (DCM) studies. Neuroimage 128, 413-431.

Friston, K.J., Preller, K.H., Mathys, C., Cagnan, H., Heinzle, J., Razi, A., Zeidman, P., 2017. Dynamic causal modelling revisited. Neuroimage.

Gecz, J., Cloosterman, D., Partington, M., 2006. ARX: a gene for all seasons. Curr Opin Genet Dev 16, 308-316.

Gelbart, M.A., Snoek, J., Adams, R.P., 2014. Bayesian Optimization with Unknown Constraints. arXiv, 1403.5607.

Gloveli, T., Dugladze, T., Saha, S., Monyer, H., Heinemann, U., Traub, R.D., Whittington, M.A., Buhl, E.H., 2005. Differential involvement of oriens/pyramidale interneurones in hippocampal network oscillations in vitro. 562, 131-147.

Hamilton, D.J., Wheeler, D.W., White, C.M., Rees, C.L., Komendantov, A.O., Bergamino, M., Ascoli, G.A., 2017. Name-calling in the hippocampus (and beyond): coming to terms with neuron types and properties. Brain Inform 4, 1-12.

Jack, J.J.B., Noble, D., Tsien, R., 1975. Electric Current Flow in Excitable Cells. Clarendon Press, Oxford.

Jansen, B.H., Rit, V.G., 1995. Electroencephalogram and visual evoked potential generation in a mathematical model of coupled cortical columns. Biol Cybern 73, 357-366.
Johnson, B.N., Palmer, C.P., Bourgeois, E.B., Elkind, J.A., Putnam, B.J., Cohen, A.S., 2014. Augmented Inhibition from Cannabinoid-Sensitive Interneurons Diminishes CA1 Output after Traumatic Brain Injury. Frontiers in Cellular Neuroscience 8.

Jung, K., Kang, J., Chung, S., Park, H.J., 2019. Dynamic causal modeling for calcium imaging: Exploration of differential effective connectivity for sensory processing in a barrel cortical column. Neuroimage 201, 116008.

Kang, J., Pae, C., Park, H.J., 2017. Energy landscape analysis of the subcortical brain network unravels system properties beneath resting state dynamics. Neuroimage 149, 153-164.

Keller, D., Ero, C., Markram, H., 2018. Cell Densities in the Mouse Brain: A Systematic Review. Front Neuroanat 12, 83.

Kiebel, S.J., Garrido, M.I., Moran, R.J., Friston, K.J., 2008. Dynamic causal modelling for EEG and MEG. Cogn Neurodyn 2, 121-136.

Krook-Magnuson, E., Armstrong, C., Oijala, M., Soltesz, I., 2013. On-demand optogenetic control of spontaneous seizures in temporal lobe epilepsy. Nat Commun 4, 1376.

Kullmann, D.M., 2011. Interneuron networks in the hippocampus. Curr Opin Neurobiol 21, 709-716.

Laurent, F., Brotons-Mas, J.R., Cid, E., Lopez-Pigozzi, D., Valero, M., Gal, B., de la Prida, L.M., 2015. Proximodistal structure of theta coordination in the dorsal hippocampus of epileptic rats. J Neurosci 35, 4760-4775.

Liley, D.T.J., Cadusch, P.J., Dafilis, M.P., 2009. A spatially continuous mean field theory of electrocortical activity. Network: Computation in Neural Systems 13, 67-113.

Lübke, J., Frotscher, M., Spruston, N., 1998. Specialized Electrophysiological Properties of Anatomically Identified Neurons in the Hilar Region of the Rat Fascia Dentata. 79, 1518-1534.

Marreiros, A.C., Daunizeau, J., Kiebel, S.J., Friston, K.J., 2008. Population dynamics: variance and the sigmoid activation function. Neuroimage 42, 147-157.

Marreiros, A.C., Kiebel, S.J., Friston, K.J., 2010. A dynamic causal model study of neuronal population dynamics. Neuroimage 51, 91-101.

Marsh, E., Fulp, C., Gomez, E., Nasrallah, I., Minarcik, J., Sudi, J., Christian, S.L., Mancini, G., Labosky, P., Dobyns, W., Brooks-Kayal, A., Golden, J.A., 2009. Targeted loss of Arx results in a developmental epilepsy mouse model and recapitulates the human phenotype in heterozygous females. Brain 132, 1563-1576.

Marsh, E.D., Nasrallah, M.P., Walsh, C., Murray, K.A., Nicole Sunnen, C., McCoy, A., Golden, J.A., 2016. Developmental interneuron subtype deficits after targeted loss of Arx. BMC Neuroscience 17.
1 Masurkar, A.V., 2018. Towards a Circuit-Level Understanding of Hippocampal CA1 Dysfunction in Alzheimer's Disease Across Anatomical Axes. Journal of Alzheimer's Disease & Parkinsonism 08.

2 McBain, C.J., Fisahn, A., 2001. Interneurons unbound. Nature Reviews Neuroscience 2, 11.

3 Moran, R., Pinotsis, D.A., Friston, K., 2013. Neural masses and fields in dynamic causal modeling. Front Comput Neurosci 7, 57.

4 Olivetti, P.R., Noebels, J.L., 2012. Interneuron, interrupted: molecular pathogenesis of ARX mutations and X-linked infantile spasms. Curr Opin Neurobiol 22, 859-865.

5 Patel, L.S., Wenzel, H.J., Schwartzkroin, P.A., 2004. Physiological and Morphological Characterization of Dentate Granule Cells in the p35 Knock-Out Mouse Hippocampus: Evidence for an Epileptic Circuit. The Journal of Neuroscience 24, 9005.

6 Rees, C.L., Moradi, K., Ascoli, G.A., 2017. Weighing the Evidence in Peters's Rule: Does Neuronal Morphology Predict Connectivity? Trends in Neurosciences 40, 63-71.

7 Rosch, R.E., Hunter, P.R., Baldeweg, T., Friston, K.J., Meyer, M.P., 2018. Calcium imaging and dynamic causal modelling reveal brain-wide changes in effective connectivity and synaptic dynamics during epileptic seizures. PLoS Comput Biol 14, e1006375.

8 Routh, B.N., Johnston, D., Harris, K., Chitwood, R.A., 2009. Anatomical and Electrophysiological Comparison of CA1 Pyramidal Neurons of the Rat and Mouse. 102, 2288-2302.

9 Roy, D.S., Kitamura, T., Okuyama, T., Ogawa, S.K., Sun, C., Obata, Y., Yoshiki, A., Tonegawa, S., 2017. Distinct Neural Circuits for the Formation and Retrieval of Episodic Memories. Cell 170, 1000-1012 e1019.

10 Sanjay, M., Neymotin, S.A., Krothapalli, S.B., 2015. Impaired dendritic inhibition leads to epileptic activity in a computer model of CA3. Hippocampus 25, 1336-1350.

11 Shahriari, B., Swersky, K., Wang, Z., Adams, R.P., Freitas, N.d., 2016. Taking the Human Out of the Loop: A Review of Bayesian Optimization. Proceedings of the IEEE 104, 148-175.

12 Snoek, J., Larochelle, H., Adams, R.P., 2012. Practical Bayesian Optimization of Machine Learning Algorithms. arXiv, 1206.2944v1202.

13 Spiegler, A., Kiebel, S.J., Atay, F.M., Knoche, T.R., 2010. Bifurcation analysis of neural mass models: Impact of extrinsic inputs and dendritic time constants. Neuroimage 52, 1041-1058.

14 Spruston, N., Lübke, J., Frotscher, M., 1997. Interneurons in the stratum lucidum of the rat hippocampus: An anatomical and electrophysiological characterization. 385, 427-440.
Tricoire, L., Pelkey, K.A., Erkkila, B.E., Jeffries, B.W., Yuan, X., McBain, C.J., 2011. A Blueprint for the Spatiotemporal Origins of Mouse Hippocampal Interneuron Diversity. The Journal of Neuroscience 31, 10948.

Wendling, F., Bartolomei, F., Bellanger, J.J., Chauvel, P., 2002. Epileptic fast activity can be explained by a model of impaired GABAergic dendritic inhibition. Eur J Neurosci 15, 1499-1508.

Wheeler, D.W., White, C.M., Rees, C.L., Komendantov, A.O., Hamilton, D.J., Ascoli, G.A., 2015. Hippocampome.org: a knowledge base of neuron types in the rodent hippocampus. Elife 4.

Wilson, H.R., Cowan, J.D., 1972. Excitatory and inhibitory interactions in localized populations of model neurons. Biophys J 12, 1-24.