DATA ARTICLE

Neurosharing: large-scale data sets (spike, LFP) recorded from the hippocampal-entorhinal system in behaving rats [version 1; peer review: 4 approved]

Kenji Mizuseki1-3, Kamran Diba2,4, Eva Pastalkova2,5, Jeff Teeters6, Anton Sirota2,7, György Buzsáki1,2,8

1NYU Neuroscience Institute, Langone Medical Center, New York University, New York, NY, USA
2Center for Molecular and Behavioral Neuroscience, Rutgers, The State University of New Jersey, Newark, NJ, USA
3Allen Institute for Brain Science, Seattle, WA, USA
4Department of Psychology, University of Wisconsin at Milwaukee, Milwaukee, WI, USA
5Janelia Farm Research Campus, Howard Hughes Medical Institute, Ashburn, VA, USA
6Redwood Center for Theoretical Neuroscience, University of California, Berkeley, Berkeley, CA, USA
7Centre for Integrative Neuroscience, University of Tübingen, Tübingen, Germany
8Center for Neural Science, New York University, New York, NY, USA

Abstract

Using silicon-based recording electrodes, we recorded neuronal activity of the dorsal hippocampus and dorsomedial entorhinal cortex from behaving rats. The entorhinal neurons were classified as principal neurons and interneurons based on monosynaptic interactions and wave-shapes. The hippocampal neurons were classified as principal neurons and interneurons based on monosynaptic interactions, wave-shapes and burstiness. The data set contains recordings from 7,736 neurons (6,100 classified as principal neurons, 1,132 as interneurons, and 504 cells that did not clearly fit into either category) obtained during 442 recording sessions from 11 rats (a total of 204.5 hours) while they were engaged in one of eight different behaviours/tasks. Both original and processed data (time stamp of spikes, spike waveforms, result of spike sorting and local field potential) are included, along with metadata of behavioural markers. Community-driven data sharing may offer cross-validation of findings, refinement of interpretations and facilitate discoveries.

This article is included in the INCF gateway.
This article is included in the Data: Use and Reuse collection.

Corresponding authors: Kenji Mizuseki (kenjim@alleninstitute.org), György Buzsáki (Gyorgy.Buzsaki@nyumc.org)

Competing interests: No competing interests were disclosed.

Grant information: The research was supported by Uehara Memorial Foundation Research Fellowship (KM), Astellas Foundation for Research on Metabolic Disorders Research Fellowship (KM), Japan Society for the Promotion of Science’s Postdoctoral Fellowship for Research Abroad (KM), NIH NS034994 (GB), NIH MHS4671 (GB), NS074015 (GB), National Science Foundation Grant S E 0542013 (GB), The Human Frontier Science Program (GB), the James S. McDonnell Foundation (GB), the Kavli Foundation (GB), General Electric, Inc. (GB), HHMI (EP), Patterson Trust (EP), National Science Foundation Grant 0855272 (JT).

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2014 Mizuseki K et al. This is an open access article distributed under the terms of the Creative Commons Attribution Licence, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Mizuseki K, Diba K, Pastalkova E et al. Neurosharing: large-scale data sets (spike, LFP) recorded from the hippocampal-entorhinal system in behaving rats [version 1; peer review: 4 approved] F1000Research 2014, 3:98 (https://doi.org/10.12688/f1000research.3895.1)

First published: 30 Apr 2014, 3:98 (https://doi.org/10.12688/f1000research.3895.1)
Introduction
The hippocampus and entorhinal cortex are essential structures for memory and spatial navigation. Position-tuned cells ('place cells') are present in CA1, CA3 and dentate gyrus regions. Grid cells, head direction cells, and border cells have been described in the dorsomedial entorhinal cortex, and are critical ingredients of navigation systems. The temporal coordination across the entorhinal cortex and hippocampus is secured by various oscillations, especially theta, gamma and sharp wave ripples.

We recorded activity of neurons in these brain regions while animals performed various tasks, such as linear track, open maze, T-maze with wheel running delay, plus maze and zigzag maze, as well as recordings during sleep in the home cage. Extensive technical descriptions of the data sets described in this document are available in several published papers.

Several questions related to memory, navigation, spike time patterns, population coding, neuronal interactions, neuronal classification, replay, sleep homeostasis and oscillations have been studied based on this dataset. However, this dataset may provide valuable information if subjected to yet further analyses. Improved spike sorting, neuron classification and more sophisticated analyses may extend and refine the initial conclusions and offer insights that were previously missed. For these reasons we provide both unprocessed (wide band) and processed versions of our data. In our experience, all methods have limitations and must undergo continuous revision. We believe that community-driven data sharing, cross-validation of all methods have limitations and must undergo continuous revision.

Material and methods
Animal surgery
All protocols were approved by the Institutional Animal Care and Use Committee of Rutgers University (protocol No. 90-042), and all experiments were performed at Rutgers University. Before surgery, one to four rats were housed in a single home cage (made of plastic; size L = 45 cm, W = 23.5 cm, H = 20 cm). Wood shavings were used as bedding and dry pellets were provided as food. The animals were housed in a temperature controlled (68°F), but not a specific pathogen free, environment under 12:12-hours light:dark cycle where light cycle was from 7AM to 7PM. After surgery, the rats were housed individually, and highly absorbent paper (Tech-board, Shepherd Specialty Papers) was used as bedding, and the animal's health was assessed daily by the experimenters.

Details of surgery and recovery procedures have been previously described in detail. Eleven Long Evans rats (male, 3–8 months old, 250–400 g) were deeply anesthetized with isoflurane (1–1.5%). In two rats (f01_m and g01_m), two silicon probes were implanted (one in each hemisphere) and targeted CA1 region. In three rats (gor01, pin01 and vvp01), two probes (32- and/or 64-site silicon probes) were implanted in the left dorsal hippocampus, targeted to CA1 and CA3 separately, and advanced over sessions and days through overlying neocortical and hippocampal tissue. The probe positions were: rat pin01: CA3; at a 35 degree angle to coronal plane, centered on 2.8 mm posterior and 2.6 mm lateral to bregma.

CA1: 26.5 degree angle to vertical, at a 35 degree angle to coronal plane, centered on 2.6 mm posterior and 2.4 mm lateral to bregma; rat vvp01: CA3: at a 26.5 degree angle to coronal plane, centered on 2.8 mm posterior and 2.6 mm lateral to bregma. CA1: 26.5 degree angle to vertical, parallel to sagittal plane, centered on 4.4 mm posterior and 2.3 mm lateral to bregma; rat gor01: CA3: at a 26.5 degree angle to coronal plane, centered on 3.1 mm posterior, and 3.0 mm lateral to bregma. CA1: 26.5 degree angle to vertical, at a 45 degree angle to coronal plane, centered on 4.9 mm posterior and 1.5 mm lateral to bregma. In four rats (ec013, ec014, ec016 and i01_m), 32- or 64-site silicon probe(s) were implanted in the right dorsal hippocampus and recorded from CA1, CA3 or dentate gyrus, and another 4-shank silicon probe was implanted in the right dorsocaudal medial entorhinal cortex. In one rat (ec012), one 4-shank silicon probe was implanted in the right dorsocaudal medial entorhinal cortex. In rat ec012, ec013 and ec016, the probe targeting the entorhinal cortex was positioned such that the different shanks recorded from different layers (4.5 mm lateral from the midline; 0.1 mm anterior to the edge of the transverse sinus at a 20–25 degree angle in the sagittal plane with the tip pointing toward the anterior direction). In rat i01_m, the EC probe had 4 shanks and was positioned such that all shanks recorded from the same layer. For the hippocampus probe in rats ec013, ec014 and ec016, the shanks were aligned parallel to the septo-temporal axis of the hippocampus (45 degrees parasagittal), positioned centrally at 3.5 mm posterior from bregma and 2.5 mm lateral from the midline.

For all silicon probes used, each shank had eight recording sites (160 µm² each site, 1–3-MΩ impedance), and intershank distance was 200 µm. Recordings sites were staggered to provide a two-dimensional arrangement (20 µm vertical separation). The individual silicon probes were attached to respective microdrives and moved independently and slowly to the target. Two stainless steel screws inserted above the cerebellum were used as indifferent (reference) and ground electrodes during recordings. At the end of the physiological recordings during the behavioural tasks, a small anodal DC current (2–5 µA, 10 s) was applied to recording sites 1 or 2 days before rats were deeply anesthetized and euthanized by perfusion with 10% formalin solution. The positions of the electrodes were confirmed histologically and reported previously in detail.

Behavioural testing
After the animals recovered from surgery (1 to 2 weeks), physiological signals were recorded during eight different types of behaviours mostly during light cycles (see Table 1).

(1) On an elevated linear track (250 cm × 7 cm), the animals were required to run back and forth to obtain water reward at both ends. In three animals (gor01, pin01, and vvp01), a similar elevated track was used (170 cm × 6.2 cm, with 22 cm × 22 cm end platforms) that was shortened to 79 or 125 cm in some trials.

(2) In the open field task, the rats chased randomly dispersed drops of water or pieces of Froot Loops (25 mg; Kellogg's) on an elevated open platform (180 cm × 180 cm, 120 cm × 120 cm or 100 cm × 200 cm).
Table 1. Behaviour descriptions.

<table>
<thead>
<tr>
<th>Behaviour</th>
<th>Behaviour subclass (Behaviour identifier)</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>elevated linear track</td>
<td>linear</td>
<td>Linear track, 250 cm × 7 cm.</td>
</tr>
<tr>
<td>elevated linear track</td>
<td>linearOne</td>
<td>Linear track (170 cm × 6.2 cm, with 22 cm × 22 cm end platforms) that was shortened to 79 or 125 cm in some trials. (Usually shortened but sometimes also lengthened). The same linear track was used in linearOne and linearTwo but at different locations in the same recording room. The center of the track was at the same position for linearOne and linearTwo, but the track was at fixed 36.9 degree angles from each other, corresponding to the diagonals of the 480 × 640 pixel camera.</td>
</tr>
<tr>
<td>elevated linear track</td>
<td>linearTwo</td>
<td>Exactly the same as linearOne but the linear track was at different locations in the same recording room. See linearOne.</td>
</tr>
<tr>
<td>open field</td>
<td>bigSquare</td>
<td>180 cm × 180 cm.</td>
</tr>
<tr>
<td>open field</td>
<td>bigSquarePlus</td>
<td>180 cm × 180 cm square open field, divided by plus shaped walls put in the center of the field.</td>
</tr>
<tr>
<td>open field</td>
<td>midSquare</td>
<td>120 cm × 120 cm.</td>
</tr>
<tr>
<td>open field</td>
<td>Open</td>
<td>100 cm × 200 cm</td>
</tr>
<tr>
<td>rewarded wheel-running task</td>
<td>wheel</td>
<td>Operant wheel running task, See Mizuseki et al., 2009.</td>
</tr>
<tr>
<td>alternation task in T-maze</td>
<td>Mwheel</td>
<td>Alternation task in T-maze (100 cm × 120 cm) with wheel running delay. See Pastalkova et al., 2008.</td>
</tr>
<tr>
<td>alternation task in T-maze</td>
<td>Tmaze</td>
<td>Alternation task in T-maze, the same as Mwheel but without delay period. There were 2.78 camera pixels/cm, which converts to 22.24 units/cm for the .whl files (8x compression of pixels).</td>
</tr>
<tr>
<td>elevated plus maze</td>
<td>plus</td>
<td>Plus maze. 100 cm × 100 cm.</td>
</tr>
<tr>
<td>zigzag maze</td>
<td>Zigzag</td>
<td>100 cm × 200 cm zigzag maze. See Royer et al., 2010.</td>
</tr>
<tr>
<td>wheel-running in home cage</td>
<td>wheel_home</td>
<td>Wheel running in home cage with free access to a wheel with no reinforcement.</td>
</tr>
<tr>
<td>sleep</td>
<td>sleep</td>
<td>Sleeping in home cage.</td>
</tr>
</tbody>
</table>

(3) In the rewarded wheel-running task, a wheel (diameter = 29 cm) was attached to a rectangular-shape box (39 cm × 39 cm × 39 cm). The rat was required to run in the wheel continuously for 10 seconds, after which time a piece of Froot Loop was dropped in the box as reinforcement. 

(4) In the alternation task in the T-maze (100 cm × 120 cm) with wheel running delay, the animal was required to run on a wheel attached to the waiting area for 10 sec, after which time the animal had access to the central arm of the T-maze, at the end of which the animal chose to turn right or left. The animal was rewarded with water if the choice was opposite to the previous one.

(5) In the elevated plus maze (100 cm × 100 cm), the rats were motivated to run to the ends of four corridors, where water was given every 30 s.

(6) In the zigzag maze (100 cm × 200 cm) with 11 corridors, the animals learned to run back and forth between two water wells; 100 µl of water was delivered at each well.

(7) In the wheel-running in home cage, a wheel (diameter = 29 cm) was attached to a rectangular-shape box (39 cm × 39 cm × 39 cm) which was used as a home cage during the experiment. Rats had free access to the wheel, and ran on the wheel with no reinforcement.

(8) In the sleeping session, the rat slept in the home cage.

For recording of behaviour (1) to (6), animals were water-scheduled for 23 hours prior to the experiment. Otherwise, both dry food and water were provided ad libitum. For tracking the position of the animals, two small light-emitting diodes, mounted above the headstage, were recorded by a digital video camera (SONY) at 30 Hz resolution.

Data collection and cell-type classification
Detailed information about the recording system and spike sorting has been previously described. Briefly, signals were amplified (1,000×), bandpass-filtered (1 Hz–5 kHz) and acquired continuously at 20 kHz (DataMax system; RC Electronics) or 32,552 Hz (NeuraLynx, MT) at 16-bit resolution. After recording, the signals were down-sampled to 1,250 Hz (DataMax system) or 1,252 Hz (NeuraLynx, MT) at 16-bit resolution. In electrophysiological recordings, positive polarity is from zero toward positive values. To maximize the detection of very slowly discharging (‘silent’) neurons, clustering was performed on concatenated files of several behavioural and sleep sessions recorded at the same electrode position on the same recording day. We made extensive use of publicly available analytical and display programs, which were developed in our laboratory (KlustaKwik, Neuroscope) available at http://sourceforge.net/projects/klustakwik/.
available at http://neuroscope.sourceforge.net/, Klusters available at http://klusters.sourceforge.net/, NDmanager available at http://ndmanager.sourceforge.net/). The latest available version at the time was used in each case. Spike sorting was performed automatically, using KlustaKwik, followed by manual adjustment of the clusters, with the help of autocorrelogram, cross-correlogram and spike wave-shape similarity matrix (Klusters software package). Because none of the existing spike sorting algorithms is completely automated, manual adjustment is necessary. This inevitably leads to some operator-dependent variability; therefore, provided clusters are not always identical to those used in our previous publications. Hippocampal principal cells and interneurons were separated based on their burstiness, waveforms and short-term monosynaptic interactions. Classification of principal neurons and interneurons of entorhinal cortical neurons was based on waveforms and short-term monosynaptic interactions, and described previously in detail. A total of 3,113 (CA1), 882 (CA3), 66 (DG), 491 (EC2), 568 (EC3) and 551 (EC5) principal neurons and 420 (CA1), 198 (CA3), 52 (DG), 85 (EC2), 215 (EC3) and 91 (EC5) interneurons were identified and included in this data set (see Table 2–Table 4).

The tip of the probe either moved spontaneously relative to the brain or was moved by the experimenter between recording days to record from potentially different sets of neurons. However, we cannot exclude the possibility that some neurons recorded on different days were identical, because spikes recorded on each day were clustered separately, though in some instances neurons were recorded over multiple days. When we moved the electrodes, we waited for at least an hour before recording in order to stabilize the position of electrodes.

**Data description**
The data are available at CRCNS.org (http://dx.doi.org/10.6080/K09G5JRZ). Details of the data collection, processing and storage.

| Table 2. Number of cells recorded. Top row: animal identifier. Left column: brain region. Brain region EC4 indicates either entorhinal cortex layer 3 or 5 (could not be determined which); region EC? indicates in entorhinal cortex, but without layer assignment. |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Brain region | ec012 | ec013 | ec014 | ec016 | f01_m | g01_m | gor01 | l01_m | j01_m | pin01 | vvp01 | total |
| EC2 | 311 | 180 | 112 | | | | | | | | 603 |
| EC3 | 201 | 362 | 177 | 116 | | | | | | | 856 |
| EC4 | | 276 | 57 | | | | | | | | 333 |
| EC5 | 110 | 416 | 68 | 154 | | | | | | | 748 |
| EC? | | | | | | | | | | 82 | 82 |
| Total EC | 311 | 1365 | 425 | 439 | | | | | | | 2622 |
| CA1 | 1185 | 1136 | 661 | 99 | 145 | 50 | 309 | 23 | 116 | 3724 |
| CA3 | 223 | 646 | | | | | | | | 1123 |
| DG | 41 | 94 | | | | | | | | 135 |
| Unknown | 39 | 3 | 90 | | | | | | | 132 |
| Total | 311 | 2853 | 1561 | 1840 | 99 | 145 | 203 | 394 | 90 | 172 | 7736 |

| Table 3. Number of principal cells. Top row: animal identifier. Left column: brain region. |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Brain region | ec012 | ec013 | ec014 | ec016 | f01_m | g01_m | gor01 | l01_m | j01_m | pin01 | vvp01 | total |
| EC2 | 248 | 146 | 97 | | | | | | | | 491 |
| EC3 | 140 | 239 | 101 | 88 | | | | | | | 568 |
| EC4 | 214 | 46 | | | | | | | | | 260 |
| EC5 | 89 | 300 | 34 | 128 | | | | | | | 551 |
| EC? | | | | | | | | | | 51 | 51 |
| Total EC | 229 | 1001 | 281 | 359 | | | | | | | 1921 |
| CA1 | 887 | 995 | 577 | 79 | 131 | 42 | 289 | 19 | 94 | 3113 |
| CA3 | 217 | 443 | | | | | | | | 882 |
| DG | 18 | 48 | | | | | | | | 66 |
| Unknown | 37 | 1 | 80 | | | | | | | 118 |
| Total | 229 | 2160 | 1276 | 1427 | 79 | 131 | 180 | 341 | 80 | 60 | 137 | 6100 |
of data into files are included with the data set, including scripts useful for processing the data\(^5\). Here, we briefly summarize the data description.

The number of cells recorded from each animal and brain region is shown in Table 2.

Most of the recorded cells were classified as principal neurons or interneurons. The number of cells classified as principal and interneuron are shown in Table 3 and Table 4.

The 8 types of behaviours (see Behavioural Testing section) were further subdivided into 14 behaviour subclasses based on minor differences (e.g. size of maze) and used as behaviour identifiers in the dataset (Table 1).

The data were obtained during 442 recording sessions. During each session the animal performed one of the 14 behaviour subclasses. The number of recording sessions and behaviour subclasses used with each animal is shown in Table 5. The description of each behaviour subclass is given in Table 1.

### Table 4. Number of interneurons. Top row: animal identifier. Left column: brain region.

<table>
<thead>
<tr>
<th>Brain region</th>
<th>ec012</th>
<th>ec013</th>
<th>ec014</th>
<th>ec016</th>
<th>f01_m</th>
<th>g01_m</th>
<th>gor01</th>
<th>i01_m</th>
<th>j01_m</th>
<th>pin01</th>
<th>vvp01</th>
<th>total</th>
</tr>
</thead>
<tbody>
<tr>
<td>EC2</td>
<td>45</td>
<td>27</td>
<td>13</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>85</td>
</tr>
<tr>
<td>EC3</td>
<td>37</td>
<td>89</td>
<td>66</td>
<td>23</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>215</td>
</tr>
<tr>
<td>EC4</td>
<td>31</td>
<td></td>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>39</td>
</tr>
<tr>
<td>EC5</td>
<td>16</td>
<td>36</td>
<td>20</td>
<td>19</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>91</td>
</tr>
<tr>
<td>EC?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>24</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>24</td>
</tr>
<tr>
<td>Total EC</td>
<td>53</td>
<td>201</td>
<td>113</td>
<td>63</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>454</td>
</tr>
<tr>
<td>CA1</td>
<td>205</td>
<td>90</td>
<td>46</td>
<td>19</td>
<td>13</td>
<td>8</td>
<td>14</td>
<td>3</td>
<td>22</td>
<td></td>
<td></td>
<td>420</td>
</tr>
<tr>
<td>CA3</td>
<td>4</td>
<td>174</td>
<td></td>
<td>14</td>
<td></td>
<td>2</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>198</td>
</tr>
<tr>
<td>DG</td>
<td>16</td>
<td>36</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>52</td>
</tr>
<tr>
<td>Unknown</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>6</td>
<td>8</td>
</tr>
<tr>
<td>Total</td>
<td>53</td>
<td>427</td>
<td>203</td>
<td>319</td>
<td>19</td>
<td>13</td>
<td>22</td>
<td>39</td>
<td>6</td>
<td>5</td>
<td>26</td>
<td>1132</td>
</tr>
</tbody>
</table>

### Table 5. Number of recording sessions. Top row: animal identifier. Left column: behaviour subclass.

<table>
<thead>
<tr>
<th>Behaviour subclass</th>
<th>ec012</th>
<th>ec013</th>
<th>ec014</th>
<th>ec016</th>
<th>f01_m</th>
<th>g01_m</th>
<th>gor01</th>
<th>i01_m</th>
<th>j01_m</th>
<th>pin01</th>
<th>vvp01</th>
<th>total</th>
</tr>
</thead>
<tbody>
<tr>
<td>bigSquare</td>
<td>24</td>
<td>45</td>
<td>4</td>
<td>13</td>
<td></td>
<td>1</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>91</td>
</tr>
<tr>
<td>bigSquarePlus</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>linear</td>
<td>18</td>
<td>90</td>
<td>2</td>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>119</td>
</tr>
<tr>
<td>linearOne</td>
<td></td>
<td></td>
<td></td>
<td>3</td>
<td></td>
<td>5</td>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>119</td>
</tr>
<tr>
<td>linearTwo</td>
<td></td>
<td></td>
<td></td>
<td>3</td>
<td></td>
<td>5</td>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>119</td>
</tr>
<tr>
<td>midSquare</td>
<td>4</td>
<td>8</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>14</td>
</tr>
<tr>
<td>Mwheel</td>
<td>28</td>
<td>16</td>
<td>8</td>
<td>14</td>
<td>8</td>
<td>7</td>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>89</td>
</tr>
<tr>
<td>Open</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>plus</td>
<td>11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>11</td>
</tr>
<tr>
<td>sleep</td>
<td>19</td>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>30</td>
<td>31</td>
</tr>
<tr>
<td>Tmaze</td>
<td></td>
<td></td>
<td></td>
<td>2</td>
<td></td>
<td>3</td>
<td>1</td>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td>6</td>
</tr>
<tr>
<td>wheel</td>
<td>40</td>
<td>8</td>
<td>9</td>
<td></td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>58</td>
</tr>
<tr>
<td>wheel_home</td>
<td></td>
<td></td>
<td></td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>ZigZag</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Total</td>
<td>70</td>
<td>208</td>
<td>50</td>
<td>59</td>
<td>8</td>
<td>7</td>
<td>9</td>
<td>9</td>
<td>4</td>
<td>3</td>
<td>15</td>
<td>442</td>
</tr>
</tbody>
</table>
numbers assigned by the spike sorting are common to all sessions within a top-level directory, and are not specific to individual sessions. Details of the file organization are provided in the document “CRCNS.org hc3 data description” which is included with the data set.

**Metadata organization**

The metadata describing the data is stored in four tables that are included with the data set. Table `cell` has information about each spike sorted cell. Table `session` has information about each experimental session. Table `epos` contains information about the position of the electrodes. And table `file` has information about the “.tar.gz” and other files that are in the data set.

**Listing 1:** Create table statements for tables: cell, session, file and epos. Fields for each of these tables are documented in the comments.

```sql
create table cell
    id integer,       -- Id used to match original row number in MatLab PyrIntMap.Map matrix
topdir string,    -- top level directory containing data
animal string,    -- name of animal
ele integer,      -- electrode number
clu integer,      -- ID # in cluster files
region string,    -- brain region
nexciting integer,   -- number of cells this cell monosynaptically excited
ninhibiting integer, -- number of cells this cell monosynaptically inhibited
exciting integer,    -- physiologically identified exciting cells based on CCG analysis
inhibiting integer,  -- physiologically identified inhibiting cells based on CCG analysis
    -- (Detailed method in Mizuseki Sirota Pastalkova and Buzsaki., 2009 Neuron paper.)
excited integer,     -- based on cross-correlogram analysis, the cell is monosynaptically excited by other cells
inhibited integer,   -- based on cross-correlogram analysis, the cell is monosynaptically inhibited by other cells
fireRate real, -- meanISI=mean(bootstrap(100, 'mean', ISI)); fireRate = SampleRate/MeanISI; ISI is interspike intervals.
totalFireRate real,  -- num of spikes divided by total recording length
cellType string      -- 'p'=pyramidal, 'i'=interneuron, 'n'=not assigned as pyramidal or interneuron
);

create table session(
    id integer,     -- matches row in original MatLab Beh matrix
topdir string,  -- directory in data set containing data (tar.gz) files
session string, -- individual session name (corresponds to name of tar.gz file having data)
behavior string, -- one of: Mwheel, Open, Tmaze, Zigzag, bigSquare, bigSquarePlus,
                -- linear, linearOne, linearTwo, midSquare, plus, sleep, wheel, wheel_home
familiarity integer, -- number of times animal has done task, 1=animal did task for first time,
                      -- 2=second time, 3=third time, 10=10 or more
duration real,   -- recording length in seconds
);

create table file(
    -- information about files in hc3 dataset
    toppdir string, -- directory in data set containing data (tar.gz) files
session string, -- individual session name (corresponds to name of tar.gz file having data)
size integer,   -- number of bytes in tar.gz file
video_type string, -- 'mpg', 'm1v' or '-' (for no video file)
video_size integer -- size of video file, or 0 if no video file
);
```

These tables are provided in CSV (comma-separated values) format, Excel format, and as tables in an SQLite database. SQLite (http://www.sqlite.org/) is a free, open source, SQL database engine available for all common operating systems. These tables are related to each other through a field (named “topdir”), which has the name of top-level directories described above and is common to all four tables. The fields in each of these tables are listed in **Listing 1**. As described in file “CRCNS.org hc3 data description” the SQLite command interface can be used with these tables to generate summary statistics from the metadata and to locate data files that satisfy particular search criteria (for example, find data for cells of a specific type from a particular brain region and experimental behaviour).
create table epos(
    -- has electrode positions for each top level directory
    -- Note, some regions do not match that in cell table.
    -- Those that differ have following meanings:
    -- DGCA3: not sure if the electrode is DG or CA3.
    -- Ctx: somewhere in the cortex (above the hippocampus)
    -- CA: somewhere in the hippocampus (do not know if it is CA1, CA3 or DG)
    topdir string, -- directory in data set containing data (tar.gz) file
    animal string, -- animal name
    e1 string, -- region for electrode 1
    e2 string, -- region for electrode 2
    e3 string, -- region for electrode 3
    e4 string, -- region for electrode 4
    -- ... (e5 through e14 fields not shown)
    e15 string, -- region for electrode 15
    e16 string -- region for electrode 16
);
Open Peer Review

Current Peer Review Status: ✔️ ✔️ ✔️ ✔️ ✔️

Version 1

Reviewer Report 07 May 2014

https://doi.org/10.5256/f1000research.4171.r4603

© 2014 Knierim J. This is an open access peer review report distributed under the terms of the Creative Commons Attribution Licence, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

James Knierim
Zanvyl Krieger Mind/Brain Institute and Department of Neuroscience, Johns Hopkins University, Baltimore, MD, USA

This data set will be a valuable resource for investigators who wish to test hypotheses about hippocampal function and interaction with entorhinal cortex at the level of single unit and LFP physiology. I have not investigated the data base carefully to ensure its utility, as I assume the authors have done so. My comments are limited to a few questions about their F1000Research article describing the data base.

1. ‘Animal surgery’ - second paragraph:

Do the authors mean “in the coronal plane?” It is not clear whether the tetrodes were angled medially or laterally in that plane, or whether the authors mean that the tetrodes were angled anteriorly or posteriorly to the coronal plane. Please clarify here and in other locations in the text how the tetrodes were angled.

2. ‘Animal surgery’ - final paragraph:

Does the data base contain histological figures? If not, are they easily identifiable and accessible from published reports? It would be very useful to ensure that the precise location of tetrodes could be made available to investigators.

3. Table 1, ‘elevated linear track’:

Shouldn't the fixed angles be 74 degrees (i.e. +/- 37 degrees)?

4. ‘Behavioural testing’ item (5):

Were the rats motivated to run the corridors in any particular order? Was there any working memory component to the trials?

Competing Interests: No competing interests were disclosed.
I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Reader Comment 14 Jul 2014
Kenji Mizuseki, Osaka City University Graduate School of Medicine, Japan

1. "Do the authors mean "in the coronal plane?" It is not clear whether the tetrodes were angled medially or laterally in that plane, or whether the authors mean that the tetrodes were angled anteriorly or posteriorly to the coronal plane. Please clarify here and in other locations in the text how the tetrodes were angled."

Indeed the description of probe implantation in the previous manuscript was not detailed enough. What we meant by ‘to the coronal plane’ was ‘a plane formed by silicone probe shanks was 35 degrees to the coronal plane, roughly along the septo-temoporal axis’. We provide accurate and complete descriptions of probe implantation in the revised manuscript.

2. "Does the data base contain histological figures? If not, are they easily identifiable and accessible from published reports? It would be very useful to ensure that the precise location of tetrodes could be made available to investigators."

The data set does not contain histological figures, but all available histological figures are reported in the previous publications. In the revised manuscript, we described the relationship between the animal identifiers in this paper and the animal identifiers in the previous publications so that one can easily identify the histology of the animal.

3. "Shouldn't the fixed angles be 74 degrees (i.e. +/- 37 degrees)?"

Yes, indeed this should be 74 degrees and we fixed it.

4. "Were the rats motivated to run the corridors in any particular order? Was there any working memory component to the trials?"

The plus maze had two open and two closed arms, and each arm had a water cup at the outer end of it. Before the session started, all the four water cups were filled with water (~20μl). Once the animal consumed water in all of the water cups, water was given in two water cups in closed or open arms depending on which type of arms the animal visited and consumed the water lastly, that is, if the animal consumed the water in the closed arm lastly water was provided in water cups in the open arms (and vice versa). Afterword, when the animal consumed the water in the two water cups, water was given in the two water cups in the opposite type of arms to encourage the animal to visit all the wells with similar probability. We added this information in the revised manuscript.

Competing Interests: No competing interests were disclosed.
Mizuseki and colleagues provide a description of 442 datasets (more than 200 hours) of hippocampal *in vivo* recordings. These datasets provide cell classification as well as the raw data, in case users wish to return to the high-sample traces and re-cluster the data on their own. The results of different analyses from the database have been published, although all analysis possibilities have not been exhausted. While this is an atypical review to write (as any suggestions on improvement to the database seem to be akin to looking a gift horse in the mouth), I am hoping the authors can discuss the ramifications of providing such an extensive database. While any database has limitations (e.g. sometimes determining the EEG units is a bit of an exercise with these data in the present format), it is perhaps more advantageous and constructive to discuss what the community can do to make the most use of it. As the authors state “Community-driven data sharing may offer cross-validation of findings, refinement of interpretations and facilitate discoveries” and the challenges are now directed to those who are interested in participating in future analyses. In hopes that this will set-forth a new era of data-sharing, I hope that the authors can discuss their open database in a manner that parallels Giorgio Ascoli’s discussion on sharing neural reconstruction files (Ascoli, 2006). It should be noted that I do not expect the authors to have comprehensive answers to each of my comments below, but it might be beneficial if they would provide some initial thoughts to seed further discussion. Some points that the authors may wish to discuss include:

1. One barrier to sharing data is “the fear of being scooped” (Ascoli, 2006). For example, scientific progress will be dramatically increased through parallel (and hopefully, collaborative) data analysis. Are the authors concerned about being “scooped”? What about a group of researchers unknowingly using the database to conduct an analysis that is also a student’s PhD project? This hypothetical student may still be acquiring the expertise to keep pace with more seasoned researchers.

2. “An often unspoken resistance to the sharing… data is born out of concern for criticisms and mistakes” (Ascoli, 2006). This is an unprecedented event that Mizuseki and colleagues have set forth, by providing a comprehensive catalog of data in an unabashed manner. The first point I want to touch upon is the level of raw exposure in releasing data. I have never met a neuroscientist who believes that anyone else could’ve conducted their analyses better than themselves (save a modest few), but to find those who believe that the authors “didn’t look for the right thing” could make up a small battalion. Mizuseki and colleagues invite critics to their doorstep. Perhaps this is more similar to posing nude as a model for artist. By placing data online, it comes with judgment and the potential to be proved wrong. What do the authors believe the convention is if others follow in their footsteps? For example, if group A shares their database after publishing their results, group B downloads and analyzes their data in new light and finds the opposite results, what is the appropriate manner in handling the situation? The self-correction aspect of science is also accelerated when data is openly-shared. It remains to be seen how situations like this should/will be handled.
3. There is also a chance for the data to be used in order to stifle or impede publication when the result is dubious. That is, should it be considered “fair-play” for a reviewer to use the same database and conduct a similar analysis with results that contradict a submitted manuscript’s results? A tactic such as this only seems appropriate in “open-review formats”.

4. “A final barrier to sharing digital reconstructions relates to the reluctance to lose or give a competitive edge” (Ascoli, 2006). The release of this immense database will surely be the stronghold of many new assistant professors who are still in the initial stages of setting up a physiology laboratory. Moreover, I can see these data being used in laboratories that are heavily analysis driven and limited in their own capacity for high-density in vivo recordings. This increase in the number of people analyzing data invites competitors for the authors as well as their neutral peers. Sheer logic dictates that the authors are not afraid of competition (otherwise, why share the data? Please note that I am a cynic and have been taught by many reviewers that “scientific altruism” is as abundant as snarks and unicorns). Why should researchers not be afraid of others using this database to compete? Is it believed that these data will be used for collaboration? What can be done to emphasize collaboration across laboratories when using the database? How is authorship handled when multiple groups use these data? If I spend my time analyzing these data, only to have it published under a “group project name” (similar to Sir William Timothy Gowers’ Polymath project), do I put it on my CV?

5. The paper explicitly states that large downloads are prohibited. Does this mean that I should not download all the data? Is it OK to use these data for a class project? If so, is it more appropriate for the professor to disseminate it to the students or should the students make their own CRCNS account? Finally, more of an afterthought but along similar lines: should there be a collaborative processing code library that should be developed and maintained in parallel with the use of these data (similar to GitHub or SourceForge)?

I absolutely do not expect the authors to have complete answers to these questions nor should they carry the sole responsibility of determining the general conventions of what constitutes use versus misuse, but I do think that it is worth hearing their general thoughts. As this is the largest and most comprehensive database of in vivo hippocampal and entorhinal physiology to become available to the general scientific community, there will be an immense ramification. Scientific replication/external validation is an immediate and positive application of this database. As I have cited Timothy Gowers above, I think it would be best to leave off with his opinion of open data-sharing and collaboration: “It feels as though this process is to normal research as driving is to pushing a car.” (http://gowers.wordpress.com/2009/02/01/questions-of-procedure/).

As a field we have the opportunity to compete or collaborate. I hope that these data facilitate cross-laboratory collaboration where two groups are reticent to share their own data. For those that are interested in embracing the collaborative spirit, the CRCNS website also has a “marketplace” section where ideas and potential collaborations can be discussed. Finally, I applaud the authors for this unprecedented act of scientific altruism. I hope this will be a platform that accelerates our understanding of the entorhinal-hippocampal circuitry through collaboration.

**Competing Interests:** No competing interests were disclosed.

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.
Kenji Mizuseki, Osaka City University Graduate School of Medicine, Japan

1. One barrier to sharing data is “the fear of being scooped” (Ascoli, 2006). For example, scientific progress will be dramatically increased through parallel (and hopefully, collaborative) data analysis. Are the authors concerned about being “scooped”? What about a group of researchers unknowingly using the database to conduct an analysis that is also a student’s PhD project? This hypothetical student may still be acquiring the expertise to keep pace with more seasoned researchers.

We absolutely agree that “the fear of being scooped” is a big concern and for this reason some people feel strong resistance against data sharing. On the other hand, this fear is not specific to data sharing; similar experiments and analysis are underway by many groups independently every day with or without knowing that that is also a student or postdoc’s project in other groups. Our tentative feeling is that, if the research group (principal investigator and his/her students and postdocs) is happy to share data publicly, we do not see any reason not to do so. If the group is not willing to share data due to the fear of being scooped, we do not see any reason to force them to share the data.

2. “An often unspoken resistance to the sharing… data is born out of concern for criticisms and mistakes” (Ascoli, 2006). This is an unprecedented event that Mizuseki and colleagues have set forth, by providing a comprehensive catalog of data in an unabashed manner. The first point I want to touch upon is the level of raw exposure in releasing data. I have never met a neuroscientist who believes that anyone else could’ve conducted their analyses better than themselves (save a modest few), but to find those who believe that the authors “didn’t look for the right thing” could make up a small battalion. Mizuseki and colleagues invite critics to their doorstep. Perhaps this is more similar to posing nude as a model for artist. By placing data online, it comes with judgment and the potential to be proved wrong. What do the authors believe the convention is if others follow in their footsteps? For example, if group A shares their database after publishing their results, group B downloads and analyzes their data in new light and finds the opposite results, what is the appropriate manner in handling the situation? The self-correction aspect of science is also accelerated when data is openly-shared. It remains to be seen how situations like this should/will be handled.

We fully agree that releasing data may be judged as “risky” by many. We have added this information to the text and cite Ascoli (2006). However, we think that this ‘self-correction’ process is a sound one and accelerates the progress in science in the long run. Since the scientific community has only a limited history of data sharing, good manner of data sharing should be formed in the community through experience.

3. There is also a chance for the data to be used in order to stifle or impede publication when the result is dubious. That is, should it be considered “fair-play” for a reviewer to use the same database and conduct a similar analysis with results that contradict a submitted manuscript’s results? A tactic such as this only seems appropriate in “open-review formats”.

Of course this should be judged case by case, but in general we think the situation the reviewer described is a fair-play and even a sound one. Tactics and politics should not be
part of scientists’ vocabulary. We need to assume that all of our colleagues are honest and they assume the same about us.

4. “A final barrier to sharing digital reconstructions relates to the reluctance to lose or give a competitive edge” (Ascoli, 2006). The release of this immense database will surely be the stronghold of many new assistant professors who are still in the initial stages of setting up a physiology laboratory. Moreover, I can see these data being used in laboratories that are heavily analysis driven and limited in their own capacity for high-density in vivo recordings. This increase in the number of people analyzing data invites competitors for the authors as well as their neutral peers. Sheer logic dictates that the authors are not afraid of competition (otherwise, why share the data? Please note that I am a cynic and have been taught by many reviewers that “scientific altruism” is as abundant as snarks and unicorns). Why should researchers not be afraid of others using this database to compete? Is it believed that these data will be used for collaboration? What can be done to emphasize collaboration across laboratories when using the database? How is authorship handled when multiple groups use these data? If I spend my time analyzing these data, only to have it published under a “group project name” (similar to Sir William Timothy Gowers’ Polymath project), do I put it on my CV?

We do not have the answers to all these questions, of course. We can talk only about our opinions. Sharing data has two compelling reasons (1) The experiments were largely performed using public money, therefore it is only natural to return our results (data) to public at some point. (2) We cannot exhaust the every single possible analysis by ourselves, and other researchers may have much better idea and capability to analyze the data set to provide new insights.

5. The paper explicitly states that large downloads are prohibited. Does this mean that I should not download all the data? Is it OK to use these data for a class project? If so, is it more appropriate for the professor to disseminate it to the students or should the students make their own CRCNS account? Finally, more of an afterthought but along similar lines: should there be a collaborative processing code library that should be developed and maintained in parallel with the use of these data (similar to GitHub or SourceForge)?

CRCNS terms allow downloading all the data. The guidelines (not preconditions) suggest that “Unnecessary downloading of large data files is not permitted.” CRCNS just wants to avoid that industry and for-profit organizations will inappropriately use public data. If someone needs to download all the files for their scientific purpose, then it’s OK. We modified the terms to allow sharing of downloaded data amongst direct colleagues working with the data or to students in a classroom. The CRCNS.org website does have forum that allows users to upload attachments that contain code or to reference code being developed at other sites, such as GitHub. Thanks for the suggestion to have a collaboratively developed code library that is more directly associated with data sets.

**Competing Interests:** No competing interests were disclosed.
Shuzo Sakata
Strathclyde Institute of Pharmacy and Biomedical Sciences, University of Strathclyde, Glasgow, UK

This is an important report for the neuroscience community regarding huge data sets recorded from the dorsal hippocampus and dorsomedial entorhinal cortex of behaving rats. Because the authors have already published a number of papers with the data sets, I have no doubt that this paper and their shared data will contribute to the further advancement of this field.

This manuscript is clearly written in details. This article can be improved by clarifying the following minor points in the section Data collection and cell-type classification:

1. Cluster quality
   It is not clear whether the quality of single units was assessed objectively, for example, by measuring isolation distance.

2. Cell-type classification
   Because now researchers including the authors can classify cell-types optogenetically, it would be better to comment on potential pitfalls of the classification procedure used here. For example, it is difficult to exclude the possibility that some of “principal” neurons can be interneurons.

**Competing Interests:** No competing interests were disclosed.

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

---

Kenji Mizuseki, Osaka City University Graduate School of Medicine, Japan

1. “It is not clear whether the quality of single units was assessed objectively, for example, by measuring isolation distance.”

   To check the quality of spike sorting, we calculated isolation distance\(^1\), an interspike interval index \(^R_{2/10}\)\(^2\) and a fraction of interspike intervals less than 2 msec\(^3\). In the revised manuscript, we explain that these values are available for each neuron so that one can choose well-isolated neurons by using these isolation quality measurements before downloading the data. We also mention that an initial quality control filtering was done in selecting cells to include in the data set: 7943 cells were detected, but only 7736 of them were included in the data set; 207 were not included because they were judged to have insufficient quality.

2. “Because now researchers including the authors can classify cell-types optogenetically, it would be better to comment on potential pitfalls of the classification procedure used here.”
For example, it is difficult to exclude the possibility that some of “principal” neurons can be interneurons.

This is an important point. Even though our classification of principal neurons and interneurons is supported by short-term spike cross-correlogram and consistent with previous reports, misclassification of pyramidal cells and interneurons is inevitable to some extent by our method. For example, some types of interneurons with wide wave shapes (e.g. somatostatin positive interneurons) might have been classified as pyramidal cells by our method. The cell-type classification method should be verified and refined by optogenetical tools in the future.

We discuss this point in the revised manuscript.

References
In this work, Mizuseki et al. provide valuable information on their large-scale data sets of multi-neuronal and local field potential (LFP) recordings from the hippocampus (dentate gyrus, CA3, and CA1) as well as the entorhinal cortex across the layers (layers 2, 3, 4, 5 etc.) of behaving rats. These data sets have originally been used for a number of their excellent studies for past ten years. Here, they fully summarized the experimental conditions and results in individual recording sessions very clearly for any users to plan to analyze the data. It will be a really useful treasure map for all hippocampus researchers.

I have one minor comment. While they showed the number of isolated neurons in each brain area of each animal (Table 2 to Table 4), it is not clear yet how many neurons, at most, were recorded simultaneously. Some people will probably want to analyze spike activity in a population of many neurons. The data would be more available for the higher-order spike analysis if the authors added additional information on the best (or first choice) data set with a sufficient number of well-isolated neurons in the same brain area.

Competing Interests: No competing interests were disclosed.

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Indeed one should be able to know the number of neurons recorded in each recording sessions before downloading the data. Following the reviewer’s advice, we added tables giving the number of spikes recorded from each neuron during each session to the metadata tables describing the data set and we provide examples in document “crcns-hc3-data-description” of how to use these to find sessions useful for the researcher’s purpose before downloading the data. In addition, we include a new set of documents (in file “crcns-hc3-cell-counts.zip”) that lists the sessions with the most cells from each region (for both all cells and well isolated units) and also lists - for all sessions - the number of cells recorded from each region (both all cells and well isolated). These documents can also be used to select sessions with the largest number of simultaneously recorded neurons from particular regions before downloading the data.
Competing Interests: No competing interests were disclosed.

The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias
- You can publish traditional articles, null/negative results, case reports, data notes and more
- The peer review process is transparent and collaborative
- Your article is indexed in PubMed after passing peer review
- Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com