

CRCNS.org hc3 data description  
Version 0.9 (Aug 20, 2013)

### Overview.

The CRCNS.org hc3 data set contains multiunit recordings from different rat hippocampal regions while the animals were performing multiple behavioral tasks. There are 7736 cells in the data set. They were recorded from 11 animals. The number of cells recorded from each animal and brain region are shown in Table 1.

**Table 1: Number of cells recorded. Top row: animal name. Left column: brain region.**

	ec012	ec013	ec014	ec016	f01_m	g01_m	gor01	i01_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				82				2622
CA1		1185	1136	661	99	145	50	309		23	116	3724
CA3		223		646			153			45	56	1123
DG		41		94								135
Unknown		39							3	90		132
Total	311	2853	1561	1840	99	145	203	394	90	68	172	7736

Most of the recorded cells were classified as principle (pyramidal) neurons or interneurons. The number of cells classified as principle and interneuron are shown in tables 2 and 3.

**Table 2: Counts of principal (pyramidal) cells. Top row: animal name. Left column: brain region.**

	ec012	ec013	ec014	ec016	f01_m	g01_m	gor01	i01_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				51				1921
CA1		887	995	577	79	131	42	289		19	94	3113
CA3		217		443			138			41	43	882
DG		18		48								66
Unknown		37							1	80		118
Total	229	2160	1276	1427	79	131	180	341	80	60	137	6100

**Table 3: Counts of interneurons. Top row: animal name. Left column: brain region.**

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

The data was obtained during 442 recording sessions. During each session the animal performed one of 14 behavioral tasks. The number of recording sessions and behavioral tasks used with each animal is shown in Table 4. The description of each task is shown in table 5.

**Table 4: Number of recording sessions. Top row: animal name. Left column: behavioral task.**

	ec012	ec013	ec014	ec016	f01_m	g01_m	gor01	i01_m	j01_m	pin01	vvp01	total
bigSquare	24	45	4	13				1	4			91
bigSquarePlus		2										2
linear	18	90	2	9								119
linearOne							3				5	8
linearTwo							3				5	8
midSquare		4	8	2								14
Mwheel	28	16	8	14	8	7		8				89
Open											3	3
plus		11										11
sleep			19	10							1	30
Tmaze							2			3	1	6
wheel		40	8	9			1					58
wheel_home				2								2
ZigZag			1									1
Total	70	208	50	59	8	7	9	9	4	3	15	442

**Table 5: Behavior descriptions.**

Behavior	Description
bigSquare	180 cm by 180 cm.
bigSquarePlus	180 cm by 180 cm, divided by plus shaped walls.
linear	linear maze, 250 cm.
linearOne	Linear maze.
LinearTwo	Linear maze.
midSquare	Square, 120 cm by 120 cm.
Mwheel	Alternation task with wheel running. See Pastalkova et al., 2008
Open	Open maze.
plus	Plus maze.
sleep	Sleeping.
Tmaze	T-maze.
wheel	Operant wheel running task, See Mizuseki et al., 2009.
wheel_home	Homecage wheel running, animal run just for fun.
Zigzag	Zigzag maze. See Royer et al., 2010 JNS.

## Data files.

The data files for each recording session are stored in separate “tar.gz” files. These files are organized into top-level directories, each of which contains data for sessions recorded on the same day using the same animal and electrode placement combination. Three example top-level directories and some of the data files within them are shown in Figure 1.

```
ec012ec.11 – top level directory
|--- ec012ec.187.tar.gz - data for individual recording session
|--- ec012ec.188.tar.gz - "
|--- ec012ec.189.tar.gz - "
ec013.53 – top level directory
|--- ec013.932.tar.gz – data for individual recording session
|--- ec013.938.tar.gz
|--- ec013.938.mpg.tar.gz – mpg.tar.gz – movie showing animal movements
|--- ec013.939.tar.gz
i01_maze04 – top level directory
|--- i01_maze04_MS.001.tar.gz – data for individual recording session
|--- i01_maze04_MS.003.tar.gz
```

**Figure 1: Example top-level directories and enclosed data files.**

Most of the .tar.gz files (those without ‘mpg’ in the name) contain recorded neural data for a session. Files with suffix “.mpg.tar.gz” contain a mpg movie showing the animal movement during the corresponding session. (e.g. file “X.mpg.tar.gz” has the movie corresponding to data in “X.tar.gz”).

## Neural data files.

An example of the files within a session (extracted from file ec012ec.187.tar.gz) is shown below in Figure 2:

ec012ec.187.clu.1	ec012ec.187.fet.4	ec012ec.187.mm.3	ec012ec.187.spk.3
ec012ec.187.clu.2	ec012ec.187.m1m2.1	ec012ec.187.mm.4	ec012ec.187.spk.4
ec012ec.187.clu.3	ec012ec.187.m1m2.2	ec012ec.187.res.1	ec012ec.187.threshold.1
ec012ec.187.clu.4	ec012ec.187.m1m2.3	ec012ec.187.res.2	ec012ec.187.threshold.2
ec012ec.187.eeg	ec012ec.187.m1m2.4	ec012ec.187.res.3	ec012ec.187.threshold.3
ec012ec.187.fet.1	ec012ec.187.m1v	ec012ec.187.res.4	ec012ec.187.threshold.4
ec012ec.187.fet.2	ec012ec.187.mm.1	ec012ec.187.spk.1	ec012ec.187.whl
ec012ec.187.fet.3	ec012ec.187.mm.2	ec012ec.187.spk.2	ec012ec.187.xml

**Figure 2: Sample data files for a recording session. Extracted from file ec012ec.187.tar.gz.**

A description of these files is provided in the documentation for the CRCNS.org, hc-2 data set, since the file formats are the same. The two documents most useful are titled “crcns.org hc2 data description” and “crcns.org hc2 processing flowchart”. The document

A fact critical for the understanding the hc3 data set, which is not mentioned in the documents for the other data sets, is that in the hc3 data set, the cells found by spike sorting are common to all sessions

within a top level directory, and are not specific to individual session. This is because data from all sessions recorded from the same animal on the same day was merged to do spike sorting; and all sessions that were merged in this way are stored in the same top level directory in the hc3 data set.

A more thorough explanation of this and the neural data files is given below.

In the experiments, data was recorded using either 4, 8, 12, or 16 electrodes (shanks). Each electrode has 8 recording sites. When possible spikes are detected (by the voltage from any recording site crossing a threshold), the time of the possible spike, and a window of data surrounding the possible spike (from all recording sites on the electrode) are stored. The time of the spike is stored in a file with extension “.res.N” and the putative spike waveforms are stored in a file with extension “.spk.N”. “N” is the electrode number, which ranges from 1 to the number of electrodes. (The example in Figure 2 has four electrodes). Later, the spike waveforms are processed to generate features that can be used for spike sorting. They are stored in a file with extension .fet.N.

To do the spike sorting, all of the .fet.N files for sessions recorded on the same day using the same animal and electrode placement combination and same electrode N are concatenated and used as input to the spike sorting process. Spike sorting is done by programs KlustaKwick (<https://github.com/klusta-team/klustakwik>) for automatic spike sorting, then by Klusters, <http://klusters.sourceforge.net/> for manual adjustment. The result of these spike sorting steps is that the spikes most likely generated by the same neuron are placed into the same category (cluster). Each cluster is assigned a cluster number, which is a non-negative integer. The cluster numbers are stored in files with extension .clu.N. Within each .clu.N file, the first line indicates the number of clusters detected in that electrode during that session, and the subsequent lines have the cluster number assigned to the sequentially detected spikes. (So, after the first line, lines in the .clu.N files have the cluster number assigned to the corresponding lines in the .res.N, .spk.N and .fet.N files).

Because the session data which are in the same top level directory are combined for spike sorting, the cluster numbers in the .clu.N files in data files within the same directory refer to the same units. Example: cluster 1 in file ec012ec.187.clu.3 (Figure 2) is associated with the same unit as cluster 1 in file ec012ec.188.clu.3; since both are from the same top level directory (“ec012ec.11”, Figure 1) and thus were recorded using the same electrode placement – animal combination; and they are from the same electrode (3).

Even though the unit numbers are consistent across sessions within a top-level directory the number of units detected within sessions (number in first line of .clu.N file) may vary slightly since some units might not generate spikes during every session. For example, the first line of file:

ec012ec.187/ec012ec.187.clu.1

contains 9, indicating that 9 clusters were present in this session. However, first line of file:

ec012ec.189/ec012ec.189.clu.1

contains 8, even though that session is in the same top level directory. The reason is that the neuron for cluster number 2 was detected in the first session, but not the second. In case it's useful for the reader, Unix commands that were used to determine this are given in Figure 3. The command lists the first line (number of clusters detected) then the count for each cluster number for both of these .clu.1 files.

```

# display cluster counts for ec012ec.187.clu.1
ec012ec.187$ head -n 1 *".clu.1"; tail -n +2 *".clu.1" | sort -n | uniq -c
9 <== (9 clusters detected)
1 0
5595 1
1 2 <== cluster number 2 detected once
10440 3
1002 4
6109 5
2432 6
147 7
63421 8

# display cluster counts for ec012ec.189.clu.1
ec012ec.187$ cd ../ec012ec.189
ec012ec.189$ head -n 1 *".clu.1"; tail -n +2 *".clu.1" | sort -n | uniq -c
8 <== (8 clusters detected)
4 0
7329 1
3 3 (cluster number 2, not detected in this session).
59 4
16 5
97 6
182 7
57455 8

```

**Figure 3: Unix command to display cluster counts**

Within each .clu.N file, cluster 0 corresponds to mechanical noise (the wave shapes do not look like neuron's spike). Cluster 1 corresponds to small, unsortable spikes. These two clusters (0 and 1) should not be used for analysis of neural data since they do not correspond to successfully sorted spikes.

The text above describes the .spk, .res, .fet and .clu files. The files with other extensions are:

- .eeg – raw data, down sampled to 1.25 KHZ, in a binary format. This contains LFP signals.
- .m1m2.N and .mm.N – auxiliary file generated by the Klusters program.
- .m1v – movie file showing movement of animal. These can be played using the freely available “VLC media player”. Not all sessions include a video file. If a video file is present, the format is either .mpg or .m1v but not both. The
- .whl – file gives position of animal during the session. Extracted from the video file.
- .xml – configuration file used with neuroscope software (<http://neuroscope.sourceforge.net>).
- .threshold – generated when detecting putative spikes.
- .led – Indicates timing of led synchronization light in movie. The following sessions (topdir/session) include this file: ec014.20/ec014.260, ec014.25/ec014.350, ec014.39/ec014.723, ec014.39/ec014.729, ec014.41/ec014.765, ec014.41/ec014.771, ec016.20/ec016.288

A more detailed description of the files is in the documents for the hc2 data set mentioned earlier.

## **Metadata files.**

There are three zip files which provide information about the hc3 data set:

- crcns-hc3-channelorder.zip
- crcns-hc3-metadata-tables.zip
- crcns-hc3-original-docs.zip

### *channelorder.zip*

The channelorder.zip file gives the relative depth in the brain of different parts of the recording electrodes for many of the sessions. A document in the hc2 data set (crcns-hc2-shank\_maps.pdf) may provide this same information for some of the sessions, in an easier to understand format.

### *crcns-hc3-original-docs.zip*

This zip file contains the original metadata files included with the data set, along with intermediate files that were derived from them during preparation of the metadata tables described in the next section. These original metadata files should NOT be used as the primary source of information about the data set. They are included because they are the most complete metadata available, even though much of the information is undocumented and potentially confusing.

### *crcns-hc3-metadata-tables.zip*

The crcns-hc3-metadata-tables.zip file is the main metadata file in the data set and for most cases, is probably the only metadata file that will be needed. It includes information about the cells, brain regions, experiment sessions and files in the data set. This information is stored in the following four data tables:

- cell – information about each spike sorted cell
- session – information about each experimental session
- epos – position of electrodes
- file – information about “.tar.gz” files that are in the data set

These tables are provided in CSV (comma-separated values) format, Excel format, and as tables in an SQLite database. The specific file names and contents are given below:

- hc3-cell.csv – cell table CSV (comma separated value) format
- hc3-session.csv – session table, CSV format
- hc3-epos.csv – epos table, CSV format
- hc3-file.csv – file table, CSV format
- hc3-tables.sql – table definitions for SQLite3 database
- hc3-tables.db – SQLite database containing tables
- hc3-tables.xlsx – Excel spreadsheet containing tables. Each table is on a separate sheet.

The fields in each of these tables is given in file “hc3-tables.sql”, which is reproduced in Figure 4. The contents of this file are used to create the tables in the SQLite database. The fields are the same as in the CSV and Excel versions of the tables.

```

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 can be found 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 identified 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,       -- behavior, one of: Mwheel, Open, Tmaze, Zigzag, bigSquare, bigSquarePlus, circle
                           -- 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
    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)
    size integer,          -- number of bytes in tar.gz file
    video_type string,     -- 'mpg', 'mlv' or '-' (for no video file)
    video_size integer,    -- size of video file, or 0 if no video file
);

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
);

```

**Figure 4: Create table statements for tables: cell, session, file and epos. Fields for each of these tables is documented by the comments.**

## Using SQLite.

SQLite (<http://www.sqlite.org/>) is a free, open source, SQL data base engine. Precompiled binary versions are available for Linux, Mac and Windows; and are very easy to install. The SQLite database version of the data tables (file hc3-tables.db) can be used to do searches that combine information from multiple tables. To combine information from multiple tables, the tables are related to each other (foreign keys) by fields “topdir” and “session”. Specifically: tables “cell” and “epos” both have field “topdir”. Tables “session” and “file” both have fields “topdir” and “session”. By setting the “where” clause of an SQL statement to specify that these fields in different tables must match, the SQL select statements can be used to combine information from multiple tables.

An example session using SQLite to get information about the hc3 data set is given below. The example uses SQL (Structured Query Language). A tutorial for SQL sufficient to allow understanding the example and creating new queries is at: [http://www.w3schools.com/sql/sql\\_intro.asp](http://www.w3schools.com/sql/sql_intro.asp).

```
# Display files in directory (Unix shell)

jt$ ls
hc3-cell.csv  hc3-epos.csv  hc3-files.csv  hc3-session.csv
hc3-tables.db  hc3-tables.sql      hc3-tables.xlsx

# Start SQLite

jt$ sqlite3 hc3-tables.db
SQLite version 3.7.10 2012-01-16 13:28:40
Enter ".help" for instructions
Enter SQL statements terminated with a ";""

# Display tables in database hc3-tables.db

sqlite> .tables
cell      epos      file      session

# Display schema for table cell

sqlite> .schema cell
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 can be found 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 identified as pyramidal or
    interneuron
);
```

```

# Display topdir, session, animal and duration
# of sessions that have at least one cell in regions CA1 and EC2
# and with the behavior type Square, familiarity > 5, and has a video file.
# Limit to first 10 records found. Order by the duration

sqlite> select distinct s.topdir, s.session, e.animal, s.duration
...>   from cell c1, cell c2, session s, epos e, file f
...>   where c1.topdir = c2.topdir and c1.topdir = s.topdir and c1.topdir=e.topdir and
s.session=f.session
...>   and c1.region='CA1' and c2.region='EC2'
...>   and (s.behavior = 'bigSquare' or s.behavior='bigSquarePlus')
...>   and s.familiarity > 5
...>   and f.video_type != '-'
...>   order by s.duration desc limit 10;
ec014.24|ec014.333|ec014|5608.653 # only one session found
sqlite>

# Perform the same search as above, but this time, comment out
# the video file restriction.

sqlite> select distinct s.topdir, s.session, e.animal, s.duration
...>   from cell c1, cell c2, session s, epos e, file f
...>   where c1.topdir = c2.topdir and c1.topdir = s.topdir and c1.topdir=e.topdir and
s.session=f.session
...>   and c1.region='CA1' and c2.region='EC2'
...>   and (s.behavior = 'bigSquare' or s.behavior='bigSquarePlus')
...>   and s.familiarity > 5
...>   -- and f.video_type != '-' # comment out video file restriction
...>   order by s.duration desc limit 10;
ec014.24|ec014.333|ec014|5608.653 # now 10 sessions are found
ec013.42|ec013.754|ec013|3234.4
ec013.36|ec013.625|ec013|2568.5
ec013.33|ec013.553|ec013|2504.294
ec013.42|ec013.756|ec013|2423.603
ec013.42|ec013.757|ec013|2418.278
ec013.40|ec013.714|ec013|2278.1
ec013.40|ec013.713|ec013|2270.413
ec013.48|ec013.858|ec013|2233.958
ec013.45|ec013.808|ec013|2209.792
sqlite>
```

The example above shows only a small part of what could be done using SQL to find data sets of interest. Any of the fields in the tables (Figure 4) can be used and combined with other fields in other tables to identify sessions and cells that are most useful for a particular purpose.