Ground truth in ultra-dense neural recording

by

Brian D. Allen

S.M., Massachusetts Institute of Technology (2010)
B.A., Northwestern University (2005)

Submitted to the Program in Media Arts and Sciences
School of Architecture + Planning
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Media Arts and Sciences

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

February 2017

© Massachusetts Institute of Technology 2017. All rights reserved

Accepted by____

Signature redacted

Program in Media Arts and Sciences
School of Architecture + Planning

December 29, 2016

Certified by____

Signature redacted

Dr. Edward S. Boyden, III
Associate Professor of Media Arts and Sciences,
Brain and Cognitive Sciences, and Biological Engineering
Thesis Supervisor

Author____

Signature redacted

Program in Media Arts and Sciences
School of Architecture + Planning

Dr. Pattie Maes
Professor of Media Arts and Sciences
Academic Head, Media Arts and Sciences Program
Ground truth in ultra-dense neural recording
Brian D. Allen

Submitted to the Program in Media Arts and Sciences
School of Architecture + Planning
on December 29, 2016, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Media Arts and Sciences

Abstract

While biologists routinely record neural activity with multi-electrodes, spike sorting-- the
process of attributing spikes to particular neurons-- remains a challenge that typically
requires human curation. Due to technical limitations, there have been very few multi-
electrode recordings done in concert with techniques such as patch clamp, which report the
"ground truth" voltage state of a single neuron in a population. Such recordings would allow
for the direct evaluation of spike sorting, which in turn could lead to further development and
refinement of spike sorting methods.

We developed a technique to establish a whole-cell or cell-attached patch recording in a
cortical neuron of an awake or lightly anesthetized head-fixed mouse, with simultaneous
extracellular recording of the same neuron and its neighbors with arrays of close-packed,
"ultra-dense," electrodes (64-256, 9 x 9µm electrodes spaced 2µm apart on a shank). Our
recordings constitute ground truth for spike sorting evaluation, and allow for the direct
evaluation and improvement of an algorithm for automatic spike sorting that benefits from
high electrode density relative to neuron packing density. Using this technique we show the
patch-triggered extracellular waveforms of neurons at a high level of granularity distributed
across cortex, and give a glimpse into the spiking activity of the network surrounding a
patched neuron in vivo.

We explore the dataset generated with this technique and discover a spike-bursting
trajectory exhibiting apparent spike-frequency adaptation. This bursting trajectory was
readily apparent in deep but not shallow cortical neurons in patch recordings, but was
somewhat obscured in extracellular recordings, where spikes from neighboring neurons may
overlap in time to contribute "noise." We show how this trajectory can be easily seen in a
high-amplitude extracellular recording, and propose how it may be accentuated in lower
amplitude recording through the use of blind source separation.

Thesis Supervisor: Dr. Edward S. Boyden, III
Title: Associate Professor of Media Arts and Sciences, Brain and Cognitive Sciences, and
Biological Engineering
The following served on the examination committee for this dissertation:

Dr. Nancy Kopell
Professor
Department of Mathematics
Boston University

Dr. Ki Ann Goosens
Assistant Professor
Department of Brain and Cognitive Sciences
Massachusetts Institute of Technology
## Contents

1 Introduction 9
   1.1 Overview 9
   1.2 Introduction and chapter summary 11

2 Integrated whole cell and close-packed recording in vivo 16
   2.1 Close-packed "ultra-dense" multi-electrode arrays 16
   2.2 Autopatching 20
   2.3 Integration of close-packed recording and autopatching 20
   2.4 Detailed methods 23
   2.5 Contribution 27

3 Results and applications of whole cell and close-packed recording in vivo 29
   3.1 Overview 29
   3.2 Library of cells 29
   3.3 Network activity and subthreshold membrane potential 34
   3.4 Intracellular dynamics: a burst spike trajectory in deep cortical neurons 35
   3.5 A strategy to accentuate intracellular dynamics extracellularly 41
   3.6 Contribution 47

4 Principles of designing interpretable optogenetic behavior experiments 50
   4.1 Abstract 50
   4.2 Introduction 50
   4.3 Cell-autonomous side effects 51
   4.4 Network-level side effects 60
   4.5 Contribution 63
A major hurdle to understanding how neural activity leads to behavior is determining which spikes originated from which neurons, the problem of spike sorting. This problem is often non-trivial because an electrode inserted into the brain will sense spikes from many neurons in its vicinity, which may have similar amplitudes depending on their distance from the electrode (Lewicki, 1998). This may be particularly problematic when the spikes of one neuron overlap in time with those of another in its vicinity, or when a neuron's spikes are highly variable in character (Einevoll et al., 2012). For example, while spike sorting relies to some extent on the consistency of the amplitude or shape of spikes generated by a given neuron, these attributes can be modulated during events such as bursts, in which a neuron spikes repeatedly during a short period of time (e.g. within 20ms of the preceding spike (Staba et al., 2002)).

Spatial oversampling is a technique that has been successfully used to aid spike sorting, in which many small electrodes sense the spikes of the same neurons from different "angles" (this strategy is reviewed in (Buzsáki, 2004)). Similar to how two ears aid in estimation of direction of a signal over one alone, spatial oversampling allows something akin to triangulation of the originating source, thereby increasing the distinguishability of spikes originating from different neurons. How much spatial oversampling is necessary for good spike sorting results? This is an unsolved question, and is important because there is an inherent tradeoff between spanning a larger volume of tissue with sparsely-packed electrodes (which may allow for the monitoring of multiple brain regions at once), and spanning a smaller volume but with higher density and perhaps greater discriminability. The
idea of this tradeoff assumes that there are a finite and in fact particularly small number of electrodes (e.g. on the order of 100 or so) that can be simultaneously employed due to economic and technological constraints (though advances in recording technologies and data storage may one day obviate the need for this tradeoff).

The process of spike sorting requires time-consuming and potentially error-prone and biased manual curation (Wood et al., 2004), but it may be possible to exploit spatial oversampling to take the human out of the loop. The desire to evaluate how massive spatial oversampling, such as occurs when there are many more signals than sources, may allow for reliable, automatic spike sorting has recently motivated the design of arrays of close-packed, or "ultra-dense," recording electrodes (Scholvin et al., 2016). These multi-electrode arrays (MEAs), for example, may span an entire mammalian cortical column while sampling densely, with more sensors (64-256 9 x 9μm electrodes set 2μm apart) than sources (neurons) in the immediate recording volume. With in vivo recordings performed using these arrays, one could spike sort utilizing data from every electrode, and then subsample electrodes to see how much performance declines, in essence creating a "dose-response" curve where number of electrodes is the dose, and spike-sorting accuracy is the response. In this way, progress may be made in determining the degree of spatial oversampling required for any given spike sorting approach.

In order for the performance of any spike sorting technique to be validated, ideally the spiking state of one or more neurons in a population would be known with near-perfect accuracy (known as "ground truth"). If this were the case, when spike sorting returned putative time-series trains of spikes for individual neurons from the extracellular recording, each of these trains could be compared to those of the ground truth recordings (see (Einevoll 2012) for a review of this approach). Then two tiers of evaluation would occur: was there a spike train returned from spike sorting that matched a ground truth spike train to some extent? If so, how many errors of omission (missing spikes in the spike-sorted train) and omission (added spikes) occurred, and how could spike sorting parameters be modified to minimize these?
1.2 Introduction and chapter summary

The purpose of this thesis work was to design a technique to record in cortex of awake or lightly-anesthetized mammals (in this case mice, who share a similar cortical architecture with humans), while simultaneously recording ground truth for a single neuron. This was in part to allow for the evaluation of spike sorting and to inform future probe and spike sorting technique design. Ground truth recording is typically very difficult and low throughput, due to the need for the recording electrode array to be very close to the ground truth neuron (e.g. < 50μm). Therefore it has rarely been done, and has never been done in awake mammals or in concert with electrodes of this density. In the past, ground truth recordings have been accomplished in anesthetized animals or slices of brain tissue through recording with a glass micropipette with a very small tip (e.g. <3μm), containing a wire and conductive solution, which either penetrates a neuron (known as intracellular or sharp recording) (Freygang and Frank, 1959; Henze et al., 2000), makes a seal with a neuron's membrane (known as cell-attached or juxtacellular patching) (Neto et al., 2016), or makes a seal and uses suction to remove a small piece of membrane (known as whole-cell patching) (Anastassiou et al., 2015). This thesis details the novel approach we developed for ground truth recording, describes the resultant dataset, and explores potential applications of the technology and dataset. This thesis does not focus on a comprehensive evaluation of spike sorting with our data, as two other theses touch on this (Bernstein, 2016; Moore-Kochlacs, 2016). However, preliminary spike sorting results are presented towards the end of chapter 3, and explored in figure 11 of that chapter.

The key insight allowing for the successful integration of close-packed electrode and whole cell recording was that a signal that was periodically emitted from the patching pipette could be sensed on the electrode array and used for estimating inter-probe distance. This allowed for ground truth recording of neurons very close to the electrode array to be accomplished with less risk of destructively ramming the pipette into the array, and provided a measure for inter-probe distance that could be used in subsequent analyses, when measurements from
actuators were not particularly reliable. This approach relied on unique features of the robotic autopatching system (Kodandaramaiah et al., 2012) and the regular geometry of electrodes on the arrays. Chapter 2 details the development of this approach.

The collection of ground truth recordings we obtained may be informative above and beyond being a test bed for evaluating and developing spike sorting and probe designs; this is explored in depth in chapter 3. For example, many of our patch recordings were done in the whole cell configuration, which grants access to subtle, subthreshold modulations of membrane potential (representing inputs to the neuron) as well as spikes (outputs). The spiking activity of the local neuronal network was simultaneously recorded (outputs, potentially serving as inputs to the patched neuron), and therefore one could analyze their potential contribution to the membrane potential of the patched neuron.

The access to small voltage deflections afforded by whole cell recording additionally makes detection and characterization of spikes during bursts feasible, when a high amount of waveform variation may translate to very small or hard to distinguish spikes in the extracellular recording (in fact, it may be unapparent from the extracellular recording if a neuron has finished bursting, or if its spikes have simply dipped into the noise floor, where it may be obscured by "noise" from spikes from other neurons (Lewicki, 1998)). Therefore, if one wanted to determine the limits of recording and spike sorting spikes from a particular neuron (such as a patched neuron in a ground truth recording session) during bursting, it would be ideal to have a sense of how the membrane potential is expected to relate to the extracellular voltage recorded by the electrode array. This would inform what sort of signal would be expected in the extracellular recording, and perhaps lead to the development of strategies to detect and spike sort burst spikes better. With such knowledge, extrapolation may be justified more broadly about the possibility of spike detection with a given extracellular recording technology, with different densities of electrode packing.

In a pioneering study many years ago, Freygang and Frank recorded the spiking activity from inside and outside the soma of a neuron simultaneously, in order to determine the
relationship between the intracellular and extracellular waveform. They designed an equivalent circuit, based on a two-compartment model, that transformed the intracellular experimental data into waveforms resembling the experimentally-observed extracellular data (Freygang and Frank, 1959). Subsequently, it was observed experimentally that such a simple model, in which the derivative of the intracellular spike waveform is proportional to the voltage of the extracellular waveform (presumed to be a consequence of the capacitance of the membrane), though it may be somewhat complicated by an interaction between the location of the electrode with respect to a neuron's morphology (Gold et al., 2006), holds particularly well during the rising phase of the spike (Anastassiou et al., 2015; Henze et al., 2000).

With this knowledge in hand, Chapter 3 closes with an exploration of a common spike bursting trajectory we discovered through whole cell recording in deep but not shallow cortex, in which time between spikes, spike shape, and spike order vary in a stereotyped fashion, and a strategy to accentuate it in extracellular recording. Though the importance of the bursting trajectory is unknown, this strategy may prove useful for recovering this and other analog waveform dynamics in extracellular recording.

Chapter 4 closes the thesis with a review article I coauthored with Annabelle Singer and Ed Boyden on designing experiments which utilize a genetically, temporally, and spatially precise method to modulate neural activity, optogenetics, to achieve maximally interpretable results in behavioral studies (Allen et al., 2015). This grew out of work earlier in my PhD in establishing causal relationships between and within neural circuits, e.g. (Bernstein et al., 2015; Chuong et al., 2014), which can be inferred with neural recording, but which are most strongly established through cycles of recording and perturbation.


Chapter 2

Integrated whole cell and close-packed recording in vivo

2.1 Close-packed "ultra-dense" multi-electrode arrays

Spike sorting is a major hurdle to understanding how neural activity leads to behavior (Lewicki, 1998). This problem is most tractable when sensors of neural spikes are small and more numerous than neurons in their recordable volume, a principle known as spatial oversampling (Fig. 1a). This fact motivated the design of "ultra-dense" multi-electrode arrays that have many close-packed sensors (MEA), with 9μm x 9μm recording sites spaced 2μm apart with 64-256 to a shank (Scholvin et al., 2016), for use in vivo (Fig. 1b).

Figure 1  Spatial oversampling with arrays of close-packed electrodes. (a) Many close-packed electrodes (top) sense activity from each neuron, demonstrating more spatial oversampling than sparser configurations (bottom). (b) An ultra-dense array of 9μm x 9μm electrodes spaced 2μm apart. Scale bar is 50μm.
While spike sorting performance may be evaluated through computational modeling, where "ground truth" is known for each neuron (see (Einevoll et al., 2012) for a review of this approach), i.e. each neuron's spike train is known with perfect fidelity, it is unclear how well current computational models capture the complexity of brain activity. Therefore, to evaluate the performance of spike sorting on ultra-dense arrays, we designed a technique to perform a series of MEA recordings with a simultaneous ground-truth recording of a single neuron, via a whole-cell or cell-attached patch, in which spikes from the same neuron could be sensed with both modalities (Fig. 2, actual data) (Allen et al., 2016). To illustrate the approach, the timing of patch-recorded spikes (Fig. 2a) is used as a trigger for 3ms snippets of extracellular data across the MEA. When these snippets are averaged, the patched neuron's waveform across space can be clearly seen, termed the neuron's "spike field" (Fig. 2b). We designed this technique for use in lightly-anesthetized and awake mice who were headfixed (Fig. 3a), so as to maximize the applicability of our results to studies in behaving and perhaps freely-moving animals. Mice were shown visual stimuli on a monitor placed in their right visual field (Fig. 3a), while patching and multi-electrode probes sharing an electrical ground (Fig. 3b), were inserted into the primary visual cortex (Fig. 3c), targeting shallow (II/III) or deep layers (V) (see Extended Methods at end of this chapter for more details).
Figure 2 A neuron's spike field is revealed through ground-truth recording. (a) Each time the patched neuron spikes, a snapshot of data is taken on the MEA and then averaged. (b) This illustrates the spatial spread of the neuron's spikes, or “spike field.”
Whole-cell or cell-attached recording would be accomplished through the use the autopatcher system, explained below, which would prove crucial for successful integration with the MEA recording system (Chapter 2.3).
2.2 Autopatching

Recording the membrane potential from a neuron is a major challenge in vivo. This motivated the design of a robotic "autopatcher," which automates key steps in patching (Kodandaramaiah et al., 2012), and which is particularly useful for the difficult job of patching onto a neuron when it cannot be seen by eye in vivo. In the past, in vivo patching has been accomplished, for example, by advancing a pipette a small distance at a time, testing the pipette's resistance, and using that resistance as a proxy for whether the pipette tip is brushing up against the insulating membrane of a neuron. The experimenter would then have a choice of attempting a patch or advancing the pipette further. The autopatcher implements these and other steps automatically. Specifically, the system generates volleys of square wave pulses of voltage (e.g. +/- 10mV) of programmable length (e.g. 2s), measuring current and allowing for the calculation of resistance, advances 2µm if no neuron is detected while outputting a voltage of 0, and then repeats those steps until a neuron is reached as indicated by sequential increases in measured resistance. Once a putative neuron is reached, a series of automated steps is executed to establish a tight seal with the neuron and allow for a "break in," in which a whole-cell recording is established granting access to the neuron's membrane voltage.

2.3 Integration of close-packed electrode recording and autopatching

While attempting co-localized recordings in vivo (the methodology of which will be discussed in detail in the "Detailed Methods" section which follows), we discovered that the volleys of square-wave voltage pulses outputted by the autopatcher could in some cases be detected across pads of the MEA, with the signal being highly localized and corresponding to when the inter-probe distance (between pipette tip and MEA) was calculated to be very small. We hypothesized that this signal could in principle be used as a "beacon" for navigation: when the amplitude on one pad or distribution of amplitudes across the regular geometry of many pads met some metric such as a threshold, perhaps that would indicate the inter-probe distance was within a given target range, say 50µm, with high precision. If this were the case,
perhaps systematic or incidental misalignments of equipment or hysteresis in the actuators, as well as subtle bending of the MEA or relaxation of the surrounding tissue, could be detected and taken into account. However, the beacon signal was difficult to detect across the MEA unless the inter-probe distance was already hypothesized to be quite small (~20 μm), so the beacon amplitude was increased from 10mV to 100mV, corresponding to roughly 15nA of current, so it could be detected across the MEA even when the inter-probe distance was comparatively large (~200μm). This allowed for monitoring of the pipette from a safe distance for the initial insertion of the pipette into brain tissue.

While conducting initial experiments with this system and trying to gain an intuition for how inter-probe distance could be determined by beacon amplitude on the MEA, we plotted the power of the pulse voltage across the MEA in the frequency band of the square wave pulses (20Hz) to see if it would fit well with a simple physical model. The details of this are described elsewhere (Bernstein, 2016), but in short consist of fitting a $1/r$ falloff curve with a noise term to the pulse power across the MEA, which could then be used to return an estimate for distance of the pipette tip to the closest MEA pad. For validation, the estimated distance could be plotted for each 2μm step of the autopatcher-driven pipette, and a linear regression would be performed to determine how linear the pipette’s path appeared to be from the model (because in practice the rigid pipette should be driven linearly by the actuator), and to estimate the residual error at each step (Fig. 4a,b). This allowed for estimation of inter-probe distance, as well as a bound on the potential error in estimation, which in practice was often quite small (~5-10μm). This system was further validated through comparison with histology, in which the patched neuron was filled with biocytin and later stained for with a fluorescent conjugate (streptavidin 488) and visualized with respect to a dye-labeled MEA probe track (Fig. 4c).
Figure 4 Co-localization and validation. (a) A patch pipette is lowered in 2 μm increments, sending test pulses each step. (b) These pulses, as seen on the MEA, are used to estimate distance from pipette tip to the closest electrode for each step, which can be fitted to a straight line for validation. (c)Histology can also be used for distance estimation, and as a further validation of the test pulse system. The neuron is filled with biocytin, which is conjugated to streptavidin-Alexa488 (green). The MEA track is labeled with Dil (magenta). Scale bar is 40 μm.

The ability to monitor inter-probe distance in real-time meant that as the pipette moved progressively closer to the MEA and a neuron was detected, a decision could be made to either a) attempt a seal if the pipette tip was within a target distance from the closest pad on the MEA, i.e. 15-50 μm, or b) remove the pipette and start the hunt at a greater depth if either a neuron was encountered that was too far from the MEA, or an unsafe inter-probe distance was reached (i.e. <15 μm). While this strategy was successful for our recordings in cortex, it is unclear how well this approach could be extended to work in tissue of highly
inhomogeneous resistivity, such as perhaps exists in brain regions with particularly tightly-packed cells or white matter tracts.

2.4 Detailed Methods

Whole cell patch recordings from 26 neurons in primary visual cortex of 20 head-fixed, lightly-anesthetized (0.5-1.2% isoflurane, n=23) or awake adult male mice (n=3), were performed simultaneously with co-localized, ultra-dense recordings with 64-256 channel multi-electrode arrays (Fig. 3). Of these recordings, 10 had a non-burst mean spike amplitude on the closest electrode of greater than 50uV (including 2 recordings from fully-awake mice) and were therefore included in a library of cells (LOC) for spike sorting evaluation (Fig. 5, next chapter). Another 6 neurons had non-burst mean amplitude of greater than 10uV but less than 50uV, and were included in an extension of the LOC for potential use in other applications, such as neuron localization. A number of cell-attached patch recordings were also performed, of which 2 recordings exhibited patch spikes of greater than 2mV and non-burst mean extracellular spikes of at least 50uV, to be included in the LOC for spike sorting evaluation. The LOC for spike sorting evaluation therefore consisted of 12 neurons (10 from whole cell and 2 cell-attached recordings).

Surgery

Male C57Bl/6 mice of 8-12 weeks of age were surgerized for head plate installation under isofluorane anesthesia (1.5-2.5%) with buprenorphine (0.1mg/kg) and meloxicam (2mg/kg) analgesics, following guidelines laid out by the Committee on Animal Care at MIT. Two stainless steel screws with attached stainless steel wires were implanted in the skull over the cerebellum and frontal cortex, to act as electrical reference and to provide extra stability for the head plate. The head plate was secured with C&B-Metabond dental cement, and the skull was sealed in with Kwikcast silicon elastomer (WPI Inc.).
Recordings commenced after 1-5 weeks. For the subset of awake recordings, mice were habituated to head-restraint over 3 days for 15, 30, and then 45 minutes with periodic condensed milk reward. On the day of or before recording, craniotomies were drilled under isoflurane anesthesia. 200-300μm diameter circular craniotomies were drilled stereotaxically either with a hand drill or with the autodriller robotic system (Pak et al., 2015). Coordinates for the MEA were [-2.8 A/P, 3.0 M/L] for motor-controlled (Thorlabs) probe insertion perpendicular to the brain surface (23 degrees to vertical) of the primary visual cortex (V1) to a final depth of 543μm for 64-channel recordings, spanning layer II/III, or 960-1000μm for 128 and 256-channel recordings, spanning layers II-V. Craniotomies for pipette insertion were located 500μm or 1000μm medial to the MEA craniotomy, for targeting layer II/III or V, respectively, with a pipette angled at 35-40 degrees from vertical. Craniotomies were periodically doused with saline or lactated Ringer’s solution to prevent dehydration, and were sealed up with Kwikcast (WPI Inc.) post-experiment.

Recording session

In preparation for a particular recording session, a mouse was initially anesthetized with isoflurane in an induction chamber and affixed by his head plate to a metal holder, with his body snugly inside a 3D-printed tube. A fluidic heating pad was placed in contact with the tube to maintain core body temperature. A cone was placed over the nose for continuous delivery of isoflurane anesthesia. A thin layer of eye lube (Puralube by Dechra or Lacri-Lube by Refresh) was placed over the eyes to maintain moisture. For anesthetized recording sessions, isoflurane was tuned to the lowest value between 0.5-1.2% that didn’t result in any motion of the animal. For an awake session, a mouse anesthetized as above was woken up by the cessation of isoflurane delivery. Electrical ground was shared between the patch and MEA recording apparatuses through a silver chloride pellet placed above the skull in saline or lactated Ringer’s solution. Electrical reference for the MEA recording was connected to either the cerebellum or cortex skull screw wire. The MEA was then inserted to its target depth, and allowed to settle for at least 10 minutes.
Patching

Patch recordings were performed with the aid of the Autopatcher robotic system as previously described. Standard intracellular solution was prepared as described in (Kodandaramaiah et al., 2016). Pipettes of 1.2mm diameter (Warner Instruments #G120F-4) were pulled to 4.5-8MOhm with either a Sutter P-97 flaming puller or a gravity puller (Narishige PC-10).

The neuron hunt portion of the patching session typically began from ~200µm from the target depth and proceeded until the autopatcher detected a candidate neuron to seal on to. If the pipette was estimated to be within ~50µm of the target portion of the MEA, a seal attempt was made. Otherwise the autopatcher program was directed to not attempt a seal, and either re-commence the neuron hunt or withdraw the pipette for another attempt. For many initial recordings, the distance estimation was made stereotaxically, aided by the amplitude of the voltage pulse signal from the pipette on the MEA. Later recordings utilized a 1/r physical model of voltage spread from the pulse volleys to aid in the decision.

When a candidate neuron was discovered by the autopatcher and deemed to be within ~50µm of the MEA, attempts were made to seal onto the neuron using the autopatcher's standard protocol. If a seal was formed but a membrane break-in was not achieved through suction pulses, the neuron was recorded from in cell-attached mode. If a break-in was achieved, membrane resistance and capacitance were estimated in voltage clamp mode, and the patching system was switched to current clamp mode, where 1s long hyperpolarizing and depolarizing current steps were injected to measure cell electrical properties and spiking threshold. A visual stimulus, below, was then played to elicit activity. If a neuron did not spike within a few minutes of the onset of the visual stimulus presentation, 50-150nA of current was injected in steady state to increase excitation. The visual stimulus would then be played again. In some cases, a non-spiking neuron that became active with current stimulation was then re-recorded from without current stimulation, when its basal activity had increased.
Visual stimulus

A small (4 x 6") computer monitor was placed at a roughly 45 degree angle in the mouse’s right visual field (Fig. 3a). Eight minutes of a visual stimulus were played of either sinusoidal drifting gratings (generated with Psychtoolbox) or a natural scene of reeds blowing in the wind (from the Chicago Motion Database, courtesy of Stephanie Palmer’s lab, University of Chicago) to elicit neural activity. A photodiode was placed in the lower left of the screen and was digitized with the patch signal, for synchronization. A typical recording session consisted of one or more eight-minute presentations in succession.

Multi-electrode array

Ultra-dense probes were designed and constructed by Jorg Scholvin at the MIT Microsystems Technology Laboratories. Recording sites were 9-10µm x 9-10µm, at a pitch of 10.5-11.5µm. Prior to experiments, recording sites were electroplated with the conductive polymer Poly(3,4-ethylenedioxythiophene) (PEDOT), in order to increase signal to noise ratio, to a target impedance of 300-600kOhm, with 0.5-1nA of constant current for 10-12s. Shorts and open circuits were assessed to determine a probe’s experimental suitability, and probes with at least ~90% good pads were selected for use. In some cases, the back of the probe was painted with 2% Dil (Invitrogen, D-282) in ethanol for subsequent tissue processing. Probes were rinsed with deionized water post-experiment, and were cleaned with ethanol or 0.25% trypsin-EDTA and isopropanol for re-use.

Data Acquisition

For a subset of recordings, the patch and MEA signals were acquired with separate acquisition systems, Multiclamp 700B and the Intan RHD2000 eval board, respectively, at 25kHz. To account for clock drift, a 25Hz sync pulse was recorded by both systems, for post-hoc temporal alignment. All but one of the 256-channel recordings were acquired with the
"Willow" (LeafLabs) direct-to-drive data acquisition system (Kinney et al., 2015) at 30kHz simultaneously with the patch signal, which was amplified as before with an Axoclamp 2B amplifier and routed to the Willow system for synchronized digitization.

Histology

Mice were perfused post-experiment with 4% paraformaldehyde in 1x phosphate-buffered saline (PBS). Brains were harvested from several mice post-experiment, of which three were chosen for subsequent reconstruction of the location of the patched neuron and dye track from the MEA. Tissue was sliced on a vibratome for coronal sections of 100μm thickness and washed in 1X PBS and 1X PBS with 0.5% Triton-X100 and 100mM Glycine (PBT). Slices were then put on a shaker overnight in a solution of PBT and 1:200 Steptavidin-Alexa488 and washed the following day in PBT and then PBS. Slices were mounted with Vectashield medium (Vector Labs) and imaged on a confocal microscope. Maximum intensity projections were made in ImageJ.

2.5 Contribution

I designed and executed the experiments under guidance from my thesis advisor, Ed Boyden. I would like to thank Jacob Bernstein, Justin Kinney, Caroline Moore-Kochlacs, Jorg Scholvin, Suhasa Kodandaramaiah, and Nikita Pak for help with developing the infrastructure for the project, as well as Ho-Jun Suk, Francisco Flores, Annabelle Singer, Giovanni Talei Franzesi, Ingrid Van Welie, Tim Blanche, members of Leaflabs, and many others in the Boyden lab for help and/or advice.

I developed the idea to use signals from the patch pipette as a beacon for localization, and ran the initial experiments to test its usefulness and tune output voltage. Jacob Bernstein modeled the beacon voltage to quantify and greatly improve localization, and performed internal validation of the system as described previously by fitting a linear regression. I then performed histological validation.
Bibliography


Chapter 3

Results and applications of integrated whole cell and close-packed recording in vivo

3.1 Overview

Recordings conducted with the techniques developed and described in the previous chapter are potentially of great use for evaluating and developing spike-sorting techniques, and may offer a unique glimpse into how an individual neuron's membrane potential relates to activity in neighboring neurons. Additionally, while analyzing the data, we discovered bursting dynamics in whole cell recordings that may be obscured when looked at from the perspective of an extracellular electrode. This chapter closes with an examination of those recordings, which includes an exploration of how the intracellular signal relates to the extracellular signal, and how it may be possible to see this relationship more clearly when recording with arrays of close-packed electrodes.

3.2 Library of cells

We termed the subset of recordings that we believe will be most useful for spike sorting evaluation, or otherwise most useful for modeling a neuron's location or morphology from extracellular data, as our "library of cells" (LOC). The LOC is the foundation for work evaluating and improving upon a spike-sorting approach based on blind source separation (Moore-Kochlacs, 2016; Bernstein, 2016), and we believe it will be an important resource for the spike sorting community more generally. Fig. 5: page 1, and Fig. 5: page 2, include recordings from twelve neurons with a non-burst extracellular absolute magnitude on the closest electrode to the neuron of at least 50μV (n=9 > 60μV and n=6 > 70μV, with n=2 from fully-awake mice). Six recordings of note are of lower amplitude, and therefore may be of less use for spike sorting (Fig. 5: page 3). However, like the other neurons in the library, they have a clear spike field,
defined as the extracellular signature extending across space of the patched neuron. Therefore they may be useful for validation and improvement of computational modeling of spike fields, i.e. determining how a neuron's location and morphology contribute to its extracellular signature. The recordings with good cell fills (n=3) in particular may prove especially useful to that end.
Figure 5 Library of cells, page 1. Patch scale bars are 50ms (x) and 10mV (y), unless otherwise noted. Red dot indicates electrode with largest mean extracellular spike.
Figure 5  Library of cells, page 2. Patch scale bars are 50ms (x) and 10mV (y), unless otherwise noted. Red dot indicates electrode with largest mean extracellular spike.
close electrode
non-burst mean = -14μV
n spikes = 97

 Patch trace

-15μV
257

neuron \(v\)

-35μV
254

neuron \(v\)

*awake

-17μV
675

neuron \(\xi\)

-32μV
233

neuron \(p\)

-12μV
233

neuron \(\tau\)

Figure 5 Library of cells, page 3. Red dot indicates electrode with largest mean extracellular spike.
3.3 Network activity and subthreshold membrane potential

Whole cell patching grants access to a neuron's subthreshold membrane potential as well as its spikes, while ultra-dense recording performed very close to the patched neuron detects spikes from that neuron and its neighbors in a spatially continuous fashion (spanning several cortical cell layers, for example) (Allen et al., 2016). A neuron's subthreshold membrane potential provides an analog readout of synaptic inputs to the neuron (Lee et al., 2006) (which may be transformed by dendritic processing (see (London and Häusser, 2005) for a review), which may originate from spikes originating from neurons in its local microcircuit. Here we illustrate a neuron in an awake mouse as seen from a whole cell recording and 100 channels of a close-packed electrode array, in which spiking in the local network appears to correlate with subthreshold potential, as well as with spiking of the patched neuron (Fig. 6). This technique therefore provides a window into how membrane potential of a neuron may relate to patterns of activity in the network in which the neuron is embedded.

Figure 6 Whole cell and network activity. A whole cell recording (below), and spike-filtered network activity on 100 close-packed electrodes in the vicinity of the whole cell recording (above).
3.4 Intracellular dynamics: a burst spike trajectory in deep cortical neurons

Neurons are often thought of as digital elements, where spikes may be modeled as discrete all-or-nothing events, but analog properties of spikes such as width and height may play an important computational role in the brain. For example, spike shape can influence the amount of neurotransmitter released, and may be modulated by spike timing or neuromodulators (Byrne and Kandel, 1996), as well as the recent conductance history of a neuron (de Polavieja, 2005). Spike shape modulation is particularly prominent during bursting, in which many spikes are fired in quick succession. Burst events have been demonstrated to reflect sensory information and trigger plasticity more reliably than single spikes (Lisman, 1997), and may increase the computational power of the brain, for example by encoding for a particularly large excitation state that was preceded by inactivity (Harris et al., 2001).

Burst spikes may be difficult to detect with common neural recording techniques. For example, spike sorting burst spikes with extracellular electrophysiology can be difficult due to amplitude and shape modulation (Harris et al., 2000), while the fast kinetics of bursting may lead to unresolvable spikes in approaches such as calcium imaging, where the sensor rise time might be 50ms or more (Chen et al., 2013). Whole cell patch clamping, on the other hand, allows burst spikes to be observed and distinguished with high fidelity, though not always non-trivially (see (Bernstein, 2016) for the method for spike extraction that we used).

Using whole cell recording, we found a common spiking trajectory during bursting in deep (target: layer V) but not shallow (target: layer II/III) cortical neurons during visual stimulus presentation, relating spike shape, timing, and ordering within a burst. Spikes in bursts of these neurons often rode on top of a wave of depolarization (Fig. 7a), resembling that seen in "intrinsic bursting" neurons described in deep cortex in the literature (Connors et al., 1982; Helmchen et al., 1999). These neurons displayed apparent spike-frequency adaptation (SFA) (Izhikevich, 2007), perhaps reflecting contrast adaptation in response to visual stimuli (Carandini and Ferster, 1997), in a stereotyped fashion. Spikes followed a trajectory marked in
part by increasing frequency up until an apparent minimum in time between spikes (or inter-
spike interval: ISI) was reached, when burst spiking often continued but at a decreasing
frequency (Fig. 7a, right). Spike shape was modulated non-linearly during this trajectory with
ISI, reflecting a dependence on spike order within a burst. To illustrate this, the amplitude-
normalized derivative of the rising phase of the spike (explained in detail below) was plotted
versus ISI on a log scale, with data points colored for spike order in burst, with grey
representing a non-burst spike, and [red, green, magenta, yellow, blue, and orange]
representing the [first, second, third, fourth, fifth, and sixth] spike in a burst (Fig. 7a, right). A
burst spike was defined as one whose preceding spike occurred less than 20ms (ISI<20ms)
before it (Staba et al., 2002). A first spike in a burst was defined as a spike whose subsequent
spike occurred within 20ms. While by inspection this criterion of 20ms generally appeared to
capture whether a neuron was bursting, perhaps because spikes occurring close to 20ms after
a previous spike were not always part of a burst, in many graphs, e.g. Fig. 7a, right and Fig. 7b,
right, the distinction between first and second spike in a burst may look artificial. For example,
in the aforementioned graphs, there is seemingly a vertical line at 20ms separating the red
data points (first burst spike) from the green data points (second burst spike), which does not
appear to necessarily reflect how the data points would naturally cluster. However, we believe
this sort of artifact will be present when picking any criterion value for a burst spike based on
a threshold for ISI (indeed we initially tried many which yielded the same problem but to a
greater degree), and the effect on further analyses appears to be minimal. A similar bursting
pattern to that of the whole cell recording in Fig. 7a was observed in a cell-attached recording
(Fig. 7b), indicating that this phenomenon was not an artifact of whole-cell recording, e.g. it
wasn't due to dialysis of intracellular solution of the pipette into the neuron.

A burst trajectory for a single neuron was defined as the line connecting the median data
point for each ordered spike in a burst (i.e. median for first spike to median for second in a
burst, etc.), where data points were plotted as a feature of spike shape (negative of the
normalized derivative of the rising phase of the spike, explained in detail later) versus ISI.
Example burst trajectories are the black lines with arrowheads in Fig. 7a, right, and Fig. 7b,
right. In order to determine if similar trajectories were shared among populations of neurons,
we compared trajectories for each shallow cortical neuron (target layer II/III) (Fig. 7c), with trajectories for each deep cortical neuron (target layer V) (Fig. 7d), with mean trajectories for each reflected by the black lines with arrows. We found a spiral-like bursting trajectory common to deep (Fig. 7d) but not shallow cortical neurons (Fig. 7c). Trajectories for each neuron were adjusted to be equal at the second spike in burst for comparison as follows: the median ISI of the second burst spike for each neuron was re-centered to zero, and the normalized derivative for each data point was further normalized to the value of the second burst spike. Burst spiking trajectories for all whole-cell recordings of deep and shallow cortical neurons are presented in Figures 8 and 9, respectively.
Figure 7 Spike bursting trajectories in-vivo. (a) Whole cell recording of a neuron in deep visual cortex. Representative trace demonstrating bursting over 5s with 500ms in red box expanded below. Each spike is labeled with its number in a burst, and color coded as such: 0 non-burst spike, [1, 2, 3, 4, 5, 6] spike in a burst. Right: Inverse negative rise time (norm. deriv.: explained in text) of each spike is plotted with respect to inter-spike interval (isi). Data points are color coded as before, and median for each color group is plotted in black. (b) Same as (a) but for a cell-attached recording. (c) Median curves as in a and b for 12 whole cell recordings from 9 mice in shallow cortex, with mean of all curves in black, reflecting no apparent common trajectory. Curves are aligned and now normalized to second spike in burst(renorm. deriv.). (d) Same as (c) but for 14 whole cell recordings from 11 mice, with mean suggesting a shared trajectory.
Figure 8 Whole cell burst trajectories for each deep cortical neuron (n=14). Negative of the normalized derivative of rising phase of patch spikes (see text for details) versus inter-spike interval (ISI) on log scale. Coloring: non-burst spike, [1st, 2nd, 3rd, 4th, 5th, 6th] spike in burst.
Note the feature plotted for spike shape is negative of the normalized derivative. This is for comparison later in this chapter to the extracellular spike amplitude, which relates to the negative of the rising portion of the intracellular spike's derivative (Henze et al., 2000; Anastassiou et al., 2015), a relationship which has been proposed to be linear (Freygang and Frank, 1959). Here the normalization is done with respect to spike amplitude. More specifically what is being plotted is the negative inverse of the 20-80% rise time (delta t). Normalizing away the intracellular spike amplitude (setting delta amplitude = 1) may seem suboptimal, given that this value is also thought to relate to the extracellular amplitude (Freygang and Frank, 1959). The decision to use this normalized value instead of simply the derivative was made to account for drift within and between patch recordings, caused by
transient differences in electrical access to the neuron's intracellular solution, which would affect spike amplitude, but less so timing. Results comparing intracellular and extracellular waveforms that follow later in Fig. 11 hold qualitatively when plotting the peak of the actual derivative as opposed to the aforementioned feature.

While spike frequency adaptation (SFA) can play a computational role in visual processing, e.g. (Peron and Gabbiani, 2009), and these in vivo spike bursting trajectories appear to be highly structured, at this stage it would be highly speculative to attribute a computational role to them based solely on their prominence in a discrete subpopulation of neurons. The remainder of this chapter therefore treats them as an example of subtle structure in patterns of spiking that can be discovered through patching, and recovered in extracellular recording when appropriate techniques are developed to do so.

3.5 A strategy to accentuate intracellular dynamics extracellularly

Can the bursting trajectory be detected in extracellular recordings? A linear relationship is predicted between extracellular spike amplitude and the negative of the derivative of the rising phase of the intracellular spike in a simple model (Freygang and Frank, 1959), and a qualitative similarity in the two has been reported in other ground truth recordings (Henze et al., 2000; Anastassiou et al., 2015). Therefore, we plotted extracellular spike amplitude versus inter-spike interval (Fig. 10, showing extracellular trajectories for deep cortical neurons). The y-axis amplitude was defined as the largest-magnitude extracellular spike amplitude on the electrode closest to the neuron, triggered from the timing (+/- 1ms) of the maximum derivative of the intracellular spike.

These trajectories bear a striking similarity to those of the intracellular burst trajectories of deep cortical neurons (Fig. 8), albeit many look more compressed in the y axis. This compression is likely due to overlapping spikes from other neurons being detected by the electrode (spikes will tend to interfere additively, as spikes are principally oriented in the same, negative direction). How can these overlapping spikes be subtracted out, so as to
accentuate the signal from the patched neuron? This would require a spike sorting technique that not only determines the spike timing of a target neuron, but that also separates the contribution of spikes from other neurons to the waveform of the target neuron’s spikes as seen on the electrode. Commonly used spike sorting techniques based on clustering fail to do so (Pillow et al., 2013).

The field of blind source separation (BSS) in machine learning addresses the question of how to uncover underlying sources when the data recorded is an unknown mixture of sources and noise, commonly known as the “cocktail party problem” (Bell and Sejnowski, 1995). For cases in which there are more sources than signals, the BSS algorithm known as independent component analysis (ICA) finds a linear transformation, or unmixing matrix, to reconstruct estimates of the underlying sources called components (see (Brown et al., 2001) for a review of the approach) by maximizing criteria related to statistical independence of the components. In extracellular recordings, this corresponds to unmixing the signals from electrode recordings, which typically contain spikes from many neurons, to recover the underlying spike trains from individual neurons (e.g. (Takahashi et al., 2002; Jackel et al., 2012; Leibig et al., 2016)). When this is done successfully (i.e. when signals are fully unmixed), effective signal to noise ratio for a target neuron is increased, and overlapping spikes from other neurons are subtracted out (Leibig et al., 2016).

ICA was applied to extracellular recordings from the 12 neurons in the library of cells for spike sorting evaluation. ICA returned as many time series spike trains (components) as there were electrodes. Each component was compared to the spike train of the ground truth neuron. Of these 12 recordings, seven yielded components matching the ground truth spike train to a varying degree (see (Moore-Kochlacs, 2016) for details). An optimum threshold was computed for each component, which minimized a combined error rate defined as (false positive spikes + false negative spikes)/(true positive spikes). Of the seven recordings with matching components, four were from deep cortical neurons, which exhibited the bursting trajectory. These recordings exhibited combined error rates of 17%, 9%, 52%, and 68%. Of
these, the first two, which happened to be recordings from awake mice, were chosen for further analysis, because of their low combined error rates.

Figure 10 Extracellular burst trajectories for each deep cortical neuron (n=14). Spike amplitude on closest electrode versus inter-spike interval (ISI) on log scale. Coloring: non-burst spike, [1st, 2nd, 3rd, 4th, 5th, 6th] spike in burst.

The first of these two neurons exhibited high-magnitude spikes (~260uV) on the electrode estimated to be closest to the patched neuron. This high-amplitude neuron's bursting trajectory is plotted as seen extracellularly (extra, Fig. 11a.i, middle, top) and intracellularly (Fig. 11a.i, middle, below), and extracellular spike amplitude is plotted against intracellular negative normalized derivative, as a linear relationship between extracellular amplitude and
intracellular derivative is predicted from a simple two compartment model (Freygang and Frank, 1959). Indeed a linear relationship is apparent ($p<.05, R^2 = .719$) (Fig. 11a,i, right).
awake recording #1

(i) signal on closest electrode (extra) vs. patch signal (intra)

(ii) ICA matching component vs. patch signal

Figure 11 (caption next page)
Figure 11 Bursting trajectories in neurons of awake mice, as detected extracellularly in spike amplitude or spike-sorted component, and as related to patch spike shape. (a) Schematic of a dual recording of a visual cortex neuron in an awake mouse (i). Right, above: time series of spike-filtered voltage, as detected from the one electrode of a 256-channel multi-electrode array (MEA) that had the maximum mean spike amplitude (extra max) matching the timing of the whole-cell recorded intracellular spikes (below, intra), is plotted. Right, top: Extracellular spike amplitude, triggered from patch spikes, is plotted with respect to isi and color coded for spike number in burst (extra), as in Figs 7, 8, and 9. Below: Negative normalized derivative of the rising portion of the patched spikes (-norm. deriv.) is plotted with respect to isi and color coded for spike number in burst (intra). Right: Spike amplitude of extra max amplitude is plotted with respect to -norm. deriv., illustrating a linear relationship (p<<.05, R^2 = 0.719). (ii) A form of blind source separation, independent component analysis (ICA) is employed on all electrodes of the MEA (ICA(extra max)) in an attempt to spike sort while preserving analog information, specifically spike height. Spike amplitude of the independent component whose spike times most closely matches those of the patch recording (matching component) are triggered by the latter and plotted with respect to isi (right). Right: These amplitudes are then plotted against -norm. deriv., mirroring the linear relationship seen in (i). (b) Same analyses as in (a) but for a neuron from a different awake mouse. The linear relationship between the spike amplitude of the matching independent component and -norm. deriv. is tighter than that of the extra amplitude with respect to -norm. deriv., with R^2 values of 0.764 and 0.47, respectively. This suggests that techniques such as ICA, which aim to determine underlying, analog signals that have been mixed together, may spike sort while preserving analog information, and may therefore improve extracellular detection of low-amplitude spikes in the bursting trajectory.

Instead of plotting the extracellular voltage of the closest electrode on the MEA versus the whole cell voltage (Fig. 11a.i), we now plot the time series of the independent component, derived from spike sorting using all electrodes on the MEA, that matches the timing of the spikes of the patched neuron (Fig. 11a.ii). When plotting the amplitude of spikes of the component versus ISI, triggered from the timing of the patch spikes (+/- 1ms), we see a similar plot as when plotting extracellular spike amplitude voltage (Fig. 11a.i), and now the linear relationship between the component versus the intracellular slope is ostensibly slightly tighter (R^2 = .784 versus .719). When performing the same set of analyses on a recording where the extracellular spike amplitude is much less in absolute magnitude (~70µV) (Fig. 11b), the relationship described above between the extracellular and intracellular waveform still is linear (p<<.05, R^2 = .47). This linear relationship greatly tightens when the amplitude of
the matching independent component, is instead plotted versus the intracellular slope ($R^2 = .47$ versus .764) (Fig. 11b.ii, right).

Because of the small number of recordings analyzed, results should not be seen as conclusive. However, these results suggest that ICA-based spike sorting on arrays of close-packed electrodes may accentuate features seen in whole cell recording, such as the deep-cortical bursting trajectory, in the extracellular space.

3.6 Contribution

I collected and analyzed the data, and designed the data analyses with Caroline Moore-Kochlacs, Jacob Bernstein, and Ed Boyden.
Bibliography


Chapter 4

Principles of designing interpretable optogenetic behavior experiments

The following is adapted from (Allen et al., 2015).

4.1 Abstract

Over the last decade, there has been much excitement about the use of optogenetic tools to test whether specific cells, regions, and projection pathways are necessary or sufficient for initiating, sustaining, or altering behavior. However, the use of such tools can result in side effects that can complicate experimental design or interpretation. The presence of optogenetic proteins in cells, the effects of heat and light, and the activity of specific ions conducted by optogenetic proteins can result in cellular side effects. At the network level, activation or silencing of defined neural populations can alter the physiology of local or distant circuits, sometimes in undesired ways. We discuss how, in order to design interpretable behavioral experiments employing optogenetics, one can understand, and control for, these potential confounds.

4.2 Introduction

Optogenetic tools allow for the precise control of the electrical activity of genetically targeted neurons by transporting specific ions into or out of cells in response to light. These tools are light sensitive proteins known as opsins, which are 7-transmembrane proteins that play photosensory or metabolic roles in species throughout the tree of life (Boyden, 2011). These opsins respond to light either by pumping ions into or out of cells (e.g., halorhodopsins pump chloride ions into archaea in response to light; bacteriorhodopsins and archaerhodopsins
pump protons out of archaea in response to light), or by opening an ion channel (e.g., channelrhodopsins lets cations such as sodium, protons, and calcium into eyespots of algae). By expressing these molecules in specific neurons, regions, or projection pathways, the targeted circuit elements can then be silenced or activated in response to light. Halorhodopsins and archaerhodopsins are commonly used for optical silencing of neural activity with light (Han and Boyden, 2007; Zhang et al., 2007a; Chow et al., 2010; Gradinaru et al., 2010; Han et al., 2011; Chuong et al., 2014). Channelrhodopsins are commonly used for optical activation of neural activity with light (Boyden et al., 2005; Nagel et al., 2005; Yizhar et al., 2011; Klapoetke et al., 2014).

These molecules have become widespread in neuroscience for the investigation of how specific neural circuit elements contribute to behavior, and are even being contemplated for therapeutic purposes (Chow and Boyden, 2013). This popularity has stemmed in part because in mammals the light-absorbing component of optogenetic tools (the chromophore all-trans retinal) is naturally present in the brain and body (Ishizuka et al., 2006). To the end of designing and interpreting behavior experiments using optogenetics, it is important to understand the side effects that these optogenetic proteins can cause in living cells, as well as the effects of heat and light on neural functions, and the biochemical activity of specific ions conducted by optogenetic proteins. Additionally, activation or silencing of defined neural populations can result in network-level side effects, for example through synaptically mediated activation of unanticipated downstream neurons. Here we discuss how these considerations can inform the design and interpretation of behavioral experiments that incorporate optogenetics as a tool.

4.3 Cell-autonomous side effects

Protein Expression
Expressing a protein in a cell can result in side effects in that cell. High levels of expression of any protein can, in principle, adversely affect cell health, and even result in cell death (Liu et al., 1999; Klein et al., 2006). Unfortunately expression levels are difficult to accurately characterize in vivo and thus the exact relationship between expression level and toxicity is
not well understood. Determining whether or not a given level of expression (e.g., as governed by gene dosage, promoter choice, and duration of expression) causes toxicity or other side effects is further complicated because such effects may depend on many other factors including species, brain region, cell type, and age of the animal. High expression of a protein may alter electrophysiology as well as cell health: in studies examining the effects of expressing opsins at high levels in mammalian HEK293 cells in vitro, changes were reported in the capacitance of the membrane (Zimmermann et al., 2008). *Drosophila* have expressed opsins under 20 copies of a conditional promoter, the upstream activating sequence (UAS), which may support a recent report of high light sensitivity of neurons in such flies for optogenetic activation (Klapoetke et al., 2014). As another example, the *N. pharaonis* halorhodopsin first assessed in neurons as an optogenetic silencer candidate (Han and Boyden, 2007; Zhang et al., 2007a) has been efficacious in multiple *C. elegans* studies (e.g., Wen et al., 2012), but in mammals appeared to form aggregates when expressed in cortical neurons (Zhang et al., 2007a; Gradinaru et al., 2008; Zhao et al., 2008). As a result, other silencers have grown in popularity, including the archaerhodopsin-class silencers (Chow et al., 2010), as well as halorhodopsins with appended trafficking-enhancement sequences (Gradinaru et al., 2010; Chuong et al., 2014). Curiously, one transgenic mouse has been created with the original *N. pharaonis* halorhodopsin in the hypocretin/orexin neurons of the mouse hypothalamus, and no abnormal morphology nor changes in cell electrical properties were noted, despite expression levels strong enough to elicit sleep upon light delivery to these cells (Tsunematsu et al., 2011). As another example, newborn neurons in the adult mouse hippocampus were reported to increase channelrhodopsin expression slowly over 3 months following neuronal differentiation, a relatively slow timecourse (Toni et al., 2008). Thus, it is possible that some cells, under specific gene expression conditions, may express optogenetic tools better or worse than other cells, even within the same species.

Opsins are often fused to fluorescent proteins to facilitate visualization in opsin-expressing cells. The type of fluorescent protein used can modulate opsin function. For example, in zebrafish, the unaltered *N. pharaonis* halorhodopsin appeared to clump when expressed with the fluorophore mCherry, but not as significantly when fused to the fluorophore YFP.
(Arrenberg et al., 2009). This differential effect appeared to persist even after appending trafficking sequences (e.g., compare Figs. 1B and S1B of the aforementioned citation). Clumping of a channelrhodopsin fused to mCherry also was apparent in Fig. 2E of (Atasoy et al., 2008), suggesting that mCherry may not be an ideal fluorophore for opsin usage. Similarly, transgenic animals expressing channelrhodopsin-2 fused to YFP exhibited greater optical excitability than did those with channelrhodopsin-2 fused to tdTomato (Madisen et al., 2012).

How then does one achieve efficacious and safe expression of an optogenetic tool in a cell type of interest? In mammals, viruses expressing opsins under cell-specific promoters in principle are useful, but many cell-specific promoters are not small enough to fit into a virus. Some commonly used promoters can target neurons (e.g., synapsin-1), or simply express pancellularly (e.g., CAG, EF1alpha) (Kügler et al., 2003; Dittgen et al., 2004; Betley and Sternson, 2011), when used in adeno-associated virus (AAV) or lentivirus, two popular methods of viral gene delivery into the mammalian brain. Complicating attempts to target expression, viruses themselves can have various tropisms, i.e. they can infect some cell types preferentially over others. For example, AAV serotype 2/1 has been reported to preferentially target inhibitory neurons, whereas lentivirus pseudotyped with a glycoprotein from the vesicular stomatitis virus (VSVg) has been reported to preferentially target excitatory neurons (Nathanson et al., 2009). Promoter-virus combinations may thus need to be optimized jointly in order to guarantee a high level of specificity.

Accordingly, many groups working with mammals have turned to transgenic mice, which can be equipped with large promoters inserted, for example, in bacterial artificial chromosome (BAC) form, to enable expression of an opsin in a specific cell class (Arenkiel et al., 2007; Borgius et al., 2010; Zhao et al., 2011). However, in addition to targeting gene expression to specific cells, such large promoters can, in principle, contain DNA sequences that modulate opsin or cellular functions. For example, the choline acetyltransferase (ChAT) gene is expressed in cholinergic neurons, and its promoter has been used to make transgenic mice expressing channelrhodopsin-2 in cholinergic neurons; this promoter, however, contains an
entire gene which happens to encode for the vesicular acetylcholine transporter (VACHT), and mice thus created exhibit altered cholinergic tone and concomitant changes in behavior (Kolisnyk et al., 2013).

A hybrid method achieves targeted gene expression taking advantage of the many hundreds of mice that express Cre recombinase transgenically under a specific promoter (via BAC transgenesis, knock-in, or other methods; see databases by the GENSAT project (Gong et al., 2003; Gong et al., 2007; Website: www.gensat.org; www.ncbi.nlm.nih.gov/projects/gensat) or Allen Mouse Brain Connectivity Atlas (Oh et al., 2014; Website: ©2014 Allen Institute for Brain Science. Allen Mouse Brain Connectivity Atlas [Internet]. Available from: http://connectivity.brain-map.org/). The opsin is then delivered via a virus that infects broad sets of cells but whose gene expression is gated by Cre recombinase expression in the target cells. The gene of interest is flanked by lox sites in a fashion so that Cre causes the deletion of a stop sequence or reversal of the gene sequence from a backward to forward direction (Atasoy et al., 2008; Kuhlman and Huang, 2008; Sohal et al., 2009). Alternatively, such Cre mice can be bred with transgenic mice whose genomes code for opsins in similar Cre-dependent expression cassettes, so that offspring selectively express the opsin in defined sets of cells (Madisen et al., 2012). When using such mice it is important to check for “leaky” or non-Cre dependent expression. For instance if Cre turns on early in development in some cells that later shut off expression, it still may have caused recombination that will persist in those cells throughout life (Betz et al., 1996; Metzger and Feil, 1999; Zhuang et al., 2005).

In Drosophila (Zhang et al., 2007b; Schroll et al., 2006) and zebrafish (Douglass et al., 2008; Baier and Scott, 2009), the GAL4/UAS expression system is commonly used because animals expressing GAL4 under a promoter can be crossed with animals with an opsin behind an UAS, enabling targeted expression via pairwise breeding. Of course, transgenic animals with cassettes bearing opsins under specific promoters can also be directly generated, e.g. in C. elegans (Nagel et al., 2005). Other expression systems - for instance the Tet off system (Zhu et al., 2009; Bundschuh et al., 2012), the Cre/lox (see above) or FLP/FRT systems (Schmitt et al., 2012; Fenno et al., 2014) - can also be used in many different species.
Optimizing the expression of opsins in the targeted cell type early in an experiment can save time and effort later on. When viruses are used, it is helpful to determine which virus produces optimal expression (in terms of volume of tissue infected, density of infected cells, and amplitude of expression in cells) empirically; one can obtain aliquots of viruses with many different promoters and serotypes (i.e., from a core facility), and test them in parallel in mice, evaluating histologically the outcome. For example, we often try multiple AAV serotypes e.g. 8 and 5, as well as 2, 9, and dj (if desired), varying the concentration by diluting virus down from its source concentration (e.g., 5x, 10x, etc. dilutions from the original concentration), in order to identify a viral dose that maximizes the number of target neurons infected, without resulting in overexpression within individual cells. When using viruses to infect cells, the proportion of target cells that express the opsin will vary from animal to animal. This variability can be harnessed for scientific discovery by using it to create a dose–response curve in which the proportion of cells infected is the dose (e.g. Aponte et al., 2011; Kim et al., 2012). Strategies to achieve a high proportion of opsin-expressing cells will vary depending on the opsin, cell type, and species; therefore we recommend trying multiple approaches in parallel as described above. Physiological validation is important, ideally in the context of the actual experiment. For example, in a study in zebrafish cardiology, investigators observed that halorhodopsin expression in cardiomyocytes varied among transgenic zebrafish. They screened for animals with strongest light responses, which occurred in about a quarter of the fish (Arrenberg et al., 2010), which facilitated later experimentation.

The time that opsins take to express can vary depending on the species, age, cell type, mode of delivery, or cell compartment desired for photoperturbation. Thus, empirical determination of expression at different time points after initiation of expression may be of use. For example it is not uncommon to perform experiments with mice four weeks after viral delivery (e.g., by AAV), but many groups wait six or more weeks for opsin expression to reach distal axons of an infected cell (Ciocchi et al., 2010; Witten et al., 2010; Felix-Ortiz et al., 2013; Kim et al., 2013; Tye et al., 2011), though in cholinergic neurons virally infected at a young age, just three weeks were reported to be effective for axonal expression (Kalmbach et al., 2012).
Given the potential side effects of expressing opsins in cells, assessing toxicity may be of utility for specific scientific or pre-clinical questions, and can be aided by analyzing cells with respect to accepted criteria or markers of toxicity for the cell type of interest (Miyashita et al., 2013; Han et al., 2009; Doroudchi et al., 2011; Gradinaru et al., 2007). Experiments to assess toxicity ideally should insure that an opsin (or its attached fluorophore) has not adversely altered gross cell morphology or function, and may include histology, electrical recordings, molecular profiling, and other strategies. To address "leaky" expression of opsins, e.g. in cell types other than the one targeted, one can assess whether the levels of expression in non-targeted cells are meaningful, for instance whether the undesired cells express meaningful amounts of protein, or respond to light stimulation (Madisen et al., 2012). Experiment design should enable comparison of results obtained from both opsin-infected and control cells or animals. Controls include tissue infected with viruses bearing a fluorescent protein with no opsin fused (though such viruses will express cytosolically, rather than near the membrane as if anchored to an opsin), or potentially an opsin with residues mutated to prevent light-activation but with otherwise normal folding and trafficking.

**Heat and Light**

Light, when absorbed by brain tissue, results in heating. The exact degree of heating for a given wavelength, duration, and power of light delivered depends on intrinsic brain properties, such as the local volume and rate of blood flow, as well as extrinsic properties, such as whether the brain is exposed to ambient temperatures after skull removal, which may encourage cooling (Kalmbach and Waters, 2012). Some reports have indicated that powers of light commonly used in optogenetics may result in heating of one degree Celsius or more, even with modest pulse durations and frequencies (Christie et al., 2013; Elias et al., 1987). Heating of cells has been shown to increase activity of specific kinds of neurons in multiple species (Moser et al., 1993; Long and Fee, 2008), although neurons equipped with different channels or receptors may undergo different kinds of changes in response to heating. Light has been shown to result in artifactual fMRI signals in the BOLD response, when delivered to the brain (Desai et al., 2011; Christie et al., 2013), perhaps also due to heating.
Light can also have non-thermal effects on cells: light-sensitive molecular pathways (opsin-mediated or otherwise) naturally found in optogenetically-targeted cells or their neighbors could potentially cause undesired effects. For example, recently in *Drosophila* and *C. elegans*, light has been shown to drive cellular signaling involving G-protein coupled receptors or other signaling pathways (Edwards et al., 2008; Liu et al., 2010; Xiang et al., 2010). Light-receptive molecules have been reported to exist in vertebrates as well (Okano et al., 1994; Blackshaw and Snyder, 1999; Koyanagi et al., 2013). Light may have other effects on experiments, such as causing photoelectric artifacts on neural recording electrodes (Han et al., 2009; Ayling et al., 2009). Light may also distract the animal, perhaps eliciting increased locomotion (Godsil and Fanselow, 2004) or necessitating a visual distractor to mask the optogenetic stimulus (Huber et al., 2008). Choosing wavelengths of light that the animal cannot see, or is less responsive to, may ameliorate this (Inada et al., 2011; Klapoetke et al., 2014), as can usage of animals that are engineered to be insensitive to light (Kocabas et al., 2012).

To address the potential effects of light and heat, optogenetic experimental design should include controls in which light is delivered in the same manner as the experimental condition but on tissue that is infected with a control virus that does not result in the neurons responding to light (e.g., with a fluorescent protein, as described above). As red light sensitive molecules come into use (Chuong et al., 2014; Klapoetke et al., 2014), in some species and contexts red light will incur less tissue absorption than blue or green, and thus incur less heating. If only a small section of tissue needs to be illuminated, the use of smaller light sources, lower light intensities, and/or shielding (e.g., in the form of beveled guide cannulas) could be used to limit light delivery to the targeted volume (Tye et al., 2011; Zorzos et al., 2010). Alternately, if a large area must be illuminated, strategies that can reduce unwanted heating include employing larger fibers (Goshen et al., 2011), fiber tips or lenses that spread light, or multiple light sources (Bernstein and Boyden, 2011). Finally, while optogenetic approaches are well suited to manipulate neural activity with millisecond precision, for long-duration experiments that do not require millisecond precision either switching opsins (e.g.,
step-function opsins (Berndt et al., 2009; Bamann et al., 2010; Yizhar et al., 2011)), or chemogenetic or pharmacogenetic approaches (i.e., using a heterologous receptor and a chemical agonist for that receptor (Shapiro et al., 2012)) might be well worth exploring.

Although not specific to optogenetics, the need for implants that go into the brain is also worth considering when designing an experiment. Devices to deliver light, like implanted optical fibers aimed at deep targets, can displace or damage brain tissue. This may be particularly crucial if regions overlying a region of interest are themselves involved in the behavior, or if scientific questions (e.g., development, dendritic spine plasticity, disease progression) are altered by the presence of an invasive element (Chuong et al., 2014). Recently developed opsins that are sensitive to red light (e.g. the channelrhodopsin ChrimsonR (Klapoetke et al., 2014) and the halorhodopsin Jaws (Chuong et al., 2014)) can alleviate these issues by reducing reliance upon implanted light sources. Because red light can more readily travel through brain tissue, light delivery devices can be placed farther from neurons expressing these red-shifted opsins, in some cases even outside the brain entirely.

Channel and Pump Activity
There are many ways in which opsin-mediated neural events may potentially differ from those mediated by natural input. Compared to electrical stimulation, many opsins conduct for longer periods of time, due to the intrinsic kinetics of the channels. Slow depolarizations resulting from optogenetic drive may engage endogenous ion channels, resulting for example in calcium entry through voltage-gated calcium channels (Zhang and Oertner, 2006). Prolonged channelrhodopsin-mediated depolarization can also result in depolarization block, a phenomenon in which pulses of light of sufficient duration inhibit action potential firing by causing sodium channel inactivation (Herman et al., 2014). Conversely, the hyperpolarization mediated by silencing opsins can engage mechanisms such as hyperpolarization-activated channels (e.g. Ih nonspecific cation channels, T-type calcium channels), resulting sometimes in “rebound” action potentials (seen in many papers employing either chloride pumps, proton pumps or electrical stimulation for hyperpolarization e.g. Madisen et al., 2012), which can be reduced by sculpting the waveform of the light so that such channels are not engaged.
All opsins exhibit some degree of desensitization during long periods of illumination, which is a function of the brightness and color of the light delivered; it is important to validate, especially for long duration light stimuli, the efficacy of the opsin throughout the experimental timecourse.

Beyond direct voltage driven effects, the specific ions conducted through optogenetic tools may alter neural functions via their interactions with signaling molecules. For example, channelrhodopsins pass multiple species of cation, chiefly protons and sodium, but also calcium and potassium (Nagel et al., 2003). Halorhodopsins and archaerhodopsins actively pump chloride into cells (Han and Boyden, 2007; Zhang et al., 2007a) and protons out of cells (Chow et al., 2010), respectively. All of these ions in principle can modulate cellular functions. Calcium has a signaling role in many aspects of cellular physiology, ranging from synaptic plasticity to apoptosis to gene expression (Berridge et al., 2000). It has been reported that channelrhodopsin-mediated calcium entry (using the molecule CatCh, which was engineered towards the goal of higher calcium flux (Kleinlogel et al., 2011)) through the plasma membrane enhances the function of voltage-gated sodium channels, and thus increases neural excitability. Halorhodopsins can pump enough chloride into cells to alter the reversal potential of GABA receptors (Raimondo et al., 2012). Channelrhodopsins expressed in Bergmann glia in cerebellar slices, upon illumination, drive glutamate release, most likely through intracellular acidification (Beppu et al., 2014). In the same preparation, light-driven proton pumps expressed in Bergmann glia could prevent ischemic brain damage, perhaps due to alkalinization of the cells and suppression of glutamate release.

Opsins may also be able to exert their voltage-dependent and ion-dependent signals in intracellular organelles, whether naturally or deliberately targeted there. Glial cells in culture or in visual cortex slices, expressing channelrhodopsin-2, release calcium from intracellular stores (Perea et al., 2014; Figueiredo et al., 2014), as evidenced by the blockade of opsin-mediated calcium signals when such stores are depleted with the drug thapsigargin. Deliberately targeting proton pumps to mitochondria of yeast or mammalian cells has been reported to result in cell survival with less reliance on oxidative phosphorylation (Hara et al.,
Thus, optogenetic tools may be utilized to study the effects of specific ions on specific physiological processes. If side effects from the ions conducted by specific opsins are a concern, one useful strategy is to use multiple opsins that result in similar voltage effects, but rely on different ionic species (Stefanik et al., 2012; Adesnik et al., 2012; Kimura et al., 2013).

4.4 Network-level side effects

Because neurons are connected – via synapses, gap junctions, and other interfaces – perturbation of the activity of a cell or set of cells can impact the activity of downstream neurons. While assessing the contribution of a cell or set of cells to network function is of course one of the key uses for optogenetics, activating or silencing a cell or set of cells may cause multiple effects, some of which are not understood or desired. Here we describe some examples of potentially unintended network side effects of the optical perturbation of a set of cells.

On short timescales, manipulating neural activity can lead to network effects like altering the balance of excitation and inhibition. For example, silencing a set of tonically active neurons may deprive downstream regions of input necessary to keep their basal activity levels within normal dynamic ranges of operation (e.g. Bao et al., 2002). Experimentally activating or silencing a set of neurons may result in downstream circuits being driven or deprived of activity in ways that do not occur endogenously. This is especially true if an entire population is being synchronously activated or silenced. Activating a set of excitatory neurons could recruit inhibitory neurons that then silence excitatory neurons, perhaps even the same ones that were excited by light (Han et al., 2009). Optogenetic silencing of excitatory neurons that drive inhibitory neurons could result in disinhibition of the neurons targeted by the inhibitory neurons (Han et al., 2011), with analogous possibilities for the case of silencing inhibitory neurons (Witten et al., 2010). While these phenomena reflect the action of circuits that exist in a normally-functioning brain, they should be taken into account when interpreting the behavioral impact of an optogenetic perturbation.
Plasticity and homeostatic adaptation can also result from perturbations of neural activity. Long-term potentiation, an increase in the synaptic efficacy between neurons, may be induced with presynaptic excitation and postsynaptic channelrhodopsin activation through as few as 20 pairings (Zhang and Oertner, 2006). Other changes—including homeostatic plasticity (Turrigiano et al., 1998; Thiagarajan et al., 2005; Borodinsky et al., 2004; Hnasko and Edwards, 2012), changes in synaptic function due to retrograde cannabinoid signaling (Wilson and Nicoll, 2001; Kreitzer and Regehr, 2001), nitric oxide signaling (Lev-Ram et al., 1995), and many other mechanisms that can cause side effects may also be engaged, since they are downstream of neural activation. As above, these phenomena may be either bugs or features in experiments, depending on how they are interpreted and whether they can reveal new findings in a well-controlled way. For example, optogenetic silencing of the CA1 region of the hippocampus for 30 minutes, can lead to recruitment of another brain region—the anterior cingulate—to help animals remember what contexts to fear, though such recruitment of the anterior cingulate does not occur with brief optogenetic silencing (Goshen et al., 2011), potentially implicating a role for network homeostatic plasticity in memory.

Even over short timescales, optogenetic perturbation can cause recruitment of undesired targets if the photostimulated neurons project to multiple regions. Optogenetic stimulation of axons projecting to one region may result in action potentials that backpropagate to the cell body and also end up driving other, undesired regions. Careful tracing of the anatomy of the cells of the targeted axons, i.e. through retrograde viral techniques, may alleviate this concern if it reveals that the neurons in question project only to the desired region (Betley et al., 2013; Stamatakis et al., 2013). When this approach is either untenable or reveals collateral axons to multiple brain regions, one can silence the cell body (e.g. pharmacologically) while stimulating the axon terminals (Stuber et al., 2011), but of course this also alters any computations that depend on the cell body. One can also infuse the target region of the projection with a neurotransmitter antagonist, to test whether the behavioral effect of optogenetic stimulation of the projections is abolished (Tye et al., 2011). This multiplexed usage of single neurons to control multiple targets has behavioral as well as physiological
consequences. For example, inhibition of serotonin neurons or Purkinje cells activity via optogenetic or chemogenetic means causes alteration of core body temperature and or blood pressure (Ray et al., 2011; Tsubota et al., 2012), revealing new functions not widely thought of as being associated with these cell types.

To monitor network effects of optogenetic manipulation, careful observation of network activity is crucial. Neural recordings can confirm that optogenetic manipulations are altering activity as intended or can reveal how that activity is altered in unexpected ways. Ideally one would record all possible effects, from changes in intracellular signaling to changes in spike output, in all cells in the nervous system (Marblestone et al., 2013). While this is not currently possible, the development of new technologies, from high-density neural recordings (Berenyi et al., 2014) to genetic reporters of activity (Chen et al., 2013) to new microscopy methods (Prevedel et al., 2014), may help us approach this ideal. In addition, optogenetic manipulations that are even more precise, from opsins that target specific cell compartments to light delivery systems that drive more naturalistic patterns of activity, may both reduce unintended network side effects and reveal how more specific or complex patterns of activity alter the network. For instance, two-photon holographic excitation of opsin-expressing neurons may support manipulation of activity with near-single cell resolution (Papagiakoumou et al., 2010; Andrasfalvy et al., 2010; Packer et al., 2012, 2015). In sum, while optogenetic tools allow for the temporally and spatially precise recruitment of well-defined sets of cells, these tools can also yield many potential confounds due to protein expression, ions conducted, heat and light, and network effects. Many of these concerns also apply to other causal technologies, such as chemogenetic inhibition of neurons. Here we have discussed both potential confounds of optogenetic methods and several approaches for addressing these confounds using currently available methods. Future developments will allow for more precise and complex manipulations and observations of neural circuits.
4.5 Contribution

I wrote this review with Annabelle Singer and Ed Boyden. Annabelle and I were co-lead authors and contributed equally to the work.
Bibliography


