Strategies for identifying exact structure of neural circuits with broad light microscopy connectivity probes

Yuriy Mishchenko

Department of Statistics, Columbia University

1255 Amsterdam Ave, New York, NY 10027

Brief Summary

Understanding of informational processes in the brain promises significant advances for communications and computing, as well as fundamental insights relevant for complex biological, ecological and social systems. Recent advances in neuroscience indicate that the collective organization of simple neurons is of primary significance for informational processes such as cognition, abstraction and thinking. However, detailed information about such organization currently may not be practically obtained. We describe a mathematical framework that will allow obtaining such information with existing light microscopy probes and genetic tools by compiling and representing large collections of neural connectivity probes as a system of mathematical constraints on the neural circuit architecture. In particular, we show that in one model organism, C. Elegans, routine, fast and complete such reconstructions for entire nervous system may be attainable already now. Such reconstructions may help understand empirically how changes in collective organization of neural circuits result in specific behaviors.

Abstract

Dissecting the structure of neural circuits in the brain is one of the central problems of neuroscience. Until present day, the only way to obtain complete and detailed reconstructions of neural circuits was thought to be the serial section Electron Microscopy, which could take decades to complete a small circuit. In this paper, we develop a mathematical framework that allows performing such reconstructions much faster and cheaper with existing light microscopy and genetic tools. In this framework, a collection of genetically targeted light probes of connectivity is prepared from different animals and then used to systematically deduce the circuit's connectivity. Each measurement is represented as mathematical constraint on the circuit architecture. Such constraints are then computationally combined to identify the detailed connectivity matrix for the probed circuit. Connectivity here is understood broadly, such as that between different identifiable neurons or identifiable classes of neurons, etc. This paradigm may be applied with connectivity probes such as ChR2-assisted circuit mapping, GRASP or transsynaptic viruses, and genetic targeting techniques such as Brainbow, MARCM/MADM or UAS/Gal4, in model organisms such as C. Elegans, Drosophila, zerbafish, mouse, etc. In particular, we demonstrate how, by using this paradigm, the wiring diagram between all neurons in C. Elegans may be reconstructed with GRASP and Brainbow and off-the-shelf light microscopy tools in the time span of one week or less. Described approach allows recovering exact connectivity matrix even if neurons may not be targeted individually in $\sim N_p \log N$ time (N_p is the number of nonzero entries and N is

the size of the connectivity matrix). For comparison, the minimal time that would be necessary to determine connectivity matrix directly by probing connections between individual neurons when one knows a-priory which pairs should be tested, e.g. with whole-cell patches, is ~ N_p.

Main Text

The problem of acquiring detailed and complete structure of neural circuits in the brain is one of the fundamental challenges of neuroscience. Detailed understanding of the structure of such circuits undoubtedly will be one of the necessary conditions for understanding how organization of simple information processing units, i.e. neurons, in the brain may lead to higher cognitive functions that we observe in animals. This point becomes ever more apparent as growing amount of evidence indicates the importance of collective behavior of neurons in neural circuits for computation in the brain[1-3]. Such collective dynamics is intimately coupled with the structure of the underlying circuits. Furthermore, detailed knowledge of neural connectivity is essential for planning, executing and properly interpreting the results of many loss-of-function and electrophysiology studies of organization of neural circuits[4, 5], as well as for analytic studies of neural circuits' structure and function[6, 7].

Ability to routinely produce detailed and complete reconstructions of neural circuits would be of great service to systems neuroscience. Even in the simplest model organisms, such as C. Elegans, ability to extract neural circuits routinely, under different conditions and in small amount of time would be invaluable. For example, C. Elegans is one of the workhorses of modern systems neuroscience with variety of interesting behavior patterns and excellent genetic and imaging tools developed for studying neural origins of behavior. By reconstructing a number of circuit instances in C. Elegans, one may be able to directly observe conserved and variable structures in its nervous system and understand their significance for behavior. Likewise, by reconstructing the circuit in different C. Elegans

mutants, one may be able to search for signatures of behavior anomalies in the circuit structure, etc.

Unfortunately, until now no satisfactory solution for this challenge exists. The only method with proven potential to produce the structure of the neural circuits down to the level of individual synapses simultaneously for all neurons is the serial section Electron Microscopy (ssEM). The only reconstruction of a complete circuit in existence was produced in this way – i.e. the circuit of about 300 neurons and 6000 synapses in C. Elegans[8-11]. Unfortunately, this technique is extremely labor intensive, slow and vulnerable to errors: the abovementioned circuit in C. Elegans took over a decade to complete, and yet its recent revision led to changes of nearly 10% of synapses affecting as many as 40% of neurons[8, 9]. Even amid significant and rapid developments in automation of ssEM analysis[12-15], the prospects for applications of ssEM in large circuits remain uncertain.

In the past few years, however, dramatic advances have occurred in the fields of fluorescent light microscopy and genetic manipulations that are now posed to change the landscape of the possibilities for neural circuit reconstructions[16-24]. Specifically, few recently described light microscopy techniques, such as mapping of circuits with light-gated ion channel ChR2[20], mapping synapses with recombinant green fluorescent protein GRASP[16], mapping circuits with modified transsynaptic rabies virus[22], etc., make it possible for fast and accurate estimation of the anatomical connectivity between different cells. The major limitation of these techniques is the small size of the libraries of distinct fluorescent markers, which limits them in the amount of information they ultimately provide. However, if coupled with genetic tools such as stochastic combinatorial gene expression used in

Brainbow[18], mitotic recombination used in mosaic analysis MARCM[17] or MADM[24], or UAS/Gal4 targeting using libraries of Gal4 lines[21], the power of these techniques may be greatly increased by allowing them to probe the same neural circuit repeatedly in different animals and accumulating information thus collected. We will show below how such information may be combined. Specifically, we will show that, by mathematically representing different such probes as constraints on the circuit's connectivity matrix, such weak measurements may be efficiently combined allowing to systematically deduce the set of possible circuit configurations and even identify its structure exactly.

In this framework we assume that the circuit's structure may be described in terms of certain units and connectivity among them in a way that would remain sufficiently invariant, or stereotypical, from animal to animal. The circuit units may be individual neurons, but also they could correspond to genetically or functionally defined neuronal populations, etc. Essential to our framework will be the assumption that such different units in the probed circuit are the same and identifiable from animal to animal. In simpler organisms, such as C. Elegans, different neurons are known to be identifiable. In larger animals, however, such identification of neurons or their populations will require novel genetic, imaging and computational techniques. The notion of connectivity here will be also understood broadly. This may be simple enumeration of the counts of distinct synapses between different units, quantification of the strengths of electric coupling via EPSP amplitude, or correlations between units' activities, etc. In all these cases, the formalism described below will be applicable. The aim of our deduction, therefore, will be the matrix

of such connection strengths between all units in the neural circuit, understood in the above "stereotypical" way.

We introduce the above paradigm by showing how combining GRASP and Brainbow genetic constructs in C. Elegans may allow reconstruction of its complete nervous system in one week or less with existing light-microscopy tools. GRASP[16] is a recently described genetic construct which uses fragments of Green Fluorescent Protein (so called split-GFP) to fluorescently mark synapses between selected cells. Specifically, two fragments of split-GFP code are made express independently in different cells. By itself, such incomplete GFP fragments do not fluoresce. However, at the location of synapses such split-GFP, if targeted to endogeneously pre- and post-synaptic proteins, may recombine across synaptic cleft and produce fluorescent puncta, thus, rendering selected synapses visible with light microscope. GRASP here plays the role of connectivity reporter: it allows estimating connectivity between given populations of cells. Brainbow[18], on the other hand, is a genetic construct which uses Cre/loxP recombination system to target stochastically a set of distinct fluorescent proteins (FP) to different cells. In one form of Brainbow, sequences of different FP in genome are flanked with inversely oriented loxP-sites so that activity of Cre-recombinase flips some of loxP-flanked sequences, thus, leading to their transcription in selected cells. Livet et al.[18] observe that such procedure typically leads to each FP producing in 50% of all cells. Here, we use Brainbow construct to stochastically express GRASP connectivity reporter, i.e. Brainbow acts as expression driver for GRASP: it allows producing variety of GRASP expression patterns within single genetic line. The latter point is important for high-throughput acquisition of large volumes of data. We also need a way

to identify cells expressing GRASP in each animal. This may be achieved by additionally introducing in genome a nuclei-targeted FP sequence in tandem with GRASP. The aim of this is to always express spit-GFP and nuclei-FP together so that GRASP expression patterns may be identified by observing which cells have nuclei-targeted fluorescence[25]. This proposed genetic construct is illustrated in Figure 1A.

For our purposes it is important that by combining Brainbow and GRASP, Cre/loxP system will lead neurons to express pre- and post-synaptic split-GFP parts of GRASP stochastically in different neurons. Whenever two such neurons form a synapse, a GRASP puncta will also be formed and may be detected with light microscope. This will allow evaluating the total count of GRASP puncta per given GRASP expression pattern. Assuming that identities of the neurons expressing GRASP in each animal may be also determined, we claim that these two pieces of information – counts of GRASP puncta and corresponding expression patterns – will allow recovery of full "stereotypical" connectivity in C. Elegans neural circuit. We should note also that another measure of connectivity may be used with GRASP, namely the total fluorescence strength of all GRASP puncta. Such total fluorescence would correspond to the size of all synapses targeted in each animal with GRASP. The formalism described below will be equally applicable in either case, although in the latter case the connectivity deduced will be in terms of the sizes of synaptic junctions between different cells, as opposed to the counts of distinct synapses in the former case. For clarity, we will talk below exclusively about GRASP puncta counts while keeping in mind applicability of developed formalism for such other kind of measurements.

To understand how complete connectivity may be extracted from such measurements, we first observe that in the described settings the count of GRASP puncta Q on average is $\mathrm{E}[Q] = N_s f^2$. Here N_s is the total count of all synapses in the entire circuit, and f=0.5 is the probability for one neuron to express GRASP in one trial. Now, let us assume that the connection between some neurons A and B is much stronger than on average for the circuit. Then, whenever neurons A and B express GRASP simultaneously, the count of GRASP puncta in such trials will be on average higher than $N_s f^2$. Likewise, whenever either neuron A or B do not express GRASP, the count of puncta will be on average lower than $N_s f^2$. In this way, the information about full connectivity in the circuit gets encoded in the fluctuations of Q in relation to the changes in GRASP expression patterns. In fact, the count of distinct synapses between any two neurons A and B may be found in the described settings simply from the triggered-averages of Q as follows

$$C_{AB} = \mathrm{E}_{\mathrm{AB}}[Q] + \mathrm{E}_{\bar{A}\bar{B}}[Q] - \mathrm{E}_{A\bar{B}}[Q] - \mathrm{E}_{\bar{A}B}[Q]. \quad (1)$$

Here $\mathrm{E}_{\mathrm{AB}}[Q], \mathrm{E}_{\bar{A}\bar{B}}[Q], \mathrm{E}_{A\bar{B}}[Q], \mathrm{E}_{\bar{A}B}[Q]$ are triggered-averages of Q given that A and B simultaneously express GRASP, neither A nor B express GRASP, A but not B expresses GRASP, and B but not A expresses GRASP, respectively. Although Eq.(1) is straightforward and computationally simple to implement, it is not practical requiring a huge number of trials before convergence (Figure 2A and C). Here we only discuss it to illustrate the principle behind encoding of information in Brainbow + GRASP.

In practice, connectivity in a circuit may be estimated from far smaller collection of measurements by employing more sophisticated data analysis. We formulate the

computational reconstruction problem for Brainbow + GRASP by observing that puncta counts Q in each trial may be viewed as linear constraints on the connectivity matrix C

$$Q(i) = \sum_{A \in PRE(i)} \sum_{B \in POST(i)} C_{AB} \,. \tag{2}$$

Here, i indicates a single trial and the summation is over the sample $PRE(i)$ of neurons expressing pre-synaptic part of GRASP and the sample $POST(i)$ of neurons expressing post-synaptic part of GRASP (Figure 1B). Eq.(2) is a simple representation of the fact that in the animal i we can observe all synapses between selected pre- and post-synaptic cells, and the information thus gained. Note that we intentionally assume no prior knowledge about the circuit structure here, in which case the total count Q is the only information that may be extracted from GRASP in each animal. If additional assumptions may be made about the circuit, such as constancy of sizes or locations of GRASP puncta for the same synapses in different animals, constraint (2) may be further strengthened.

We also observe that the connectivity matrix is sparse – e.g. in C. Elegans out of total 80,000 possible connections only about 2,000 are actually realized[8]. Sparseness of the connectivity matrix is a powerful additional constraint. In particular, it can be shown that almost any sparse matrix may be found *exactly* from only $\approx N_p \log N$ constraints (2) by minimizing its l_1 norm[26]. Here, N_p is the number of nonzero entries in the connectivity matrix, N is the total number of neurons, and l_1 norm is $\|C\|_{l_1} = \sum_{A=1}^{N} \sum_{B=1}^{N} |C_{AB}|$ - in our case this is simply the total count of synapses in the circuit. In other words, the solution for the following constrained optimization problem,

$$\min \sum_{A=1}^{N}\sum_{B=1}^{N}\left|C_{AB}\right|, \text{ subject to} \tag{3a}$$

$$\sum_{A=1}^{N}\sum_{B=1}^{N}\alpha_A(i)\beta_B(i)C_{AB} = Q(i),\ i=1...M\ , \tag{3b}$$

$$\alpha_A(i)=\begin{cases}1, A\in PRE(i)\\ 0, A\notin PRE(i)\end{cases}, \tag{3c}$$

$$\beta_B(i)=\begin{cases}1, B\in POST(i)\\ 0, B\notin POST(i)\end{cases}, \tag{3d}$$

will continuously converge to the exact connectivity matrix C as the number of available measurements M is increased until the exact matrix C is recovered when $M \approx N_p \log N$. $\alpha_A(i)$ and $\beta_B(i)$ here are the indicator functions for the sets *PRE*(i) and *POST*(i), and describe GRASP expression patterns in different animals. By construction, these are determined independently in each animal, e.g. from nuclei-targeted fluorescence. Optimization problem (3) is tractable and is a standard linear-programming problem[27]. Alternatively, if the total number of synapses N_s is known in advance (and in our case $N_s = \mathrm{E}[Q]/f^2$), the solution to (3) may be also obtained as the intersection of two convex sets – $\|C\|_{l_1} = N_s$ and the hyperplane of linear constraints (3b)[28]. Such intersection of convex sets may be found efficiently for problems of very large size, e.g. with the method of alternate projections[29]. Important feature of such deduction is that it may be performed from constraints (3b) of almost any form, i.e. it is absolutely not required here that the sets *PRE*(i) and *POST*(i) are produced stochastically.

According to the above analysis, complete circuit in C. Elegans should be possible to recover from $M \approx N_p \log N \approx 10{,}000 - 20{,}000$ Brainbow + GRASP measurements. To test whether such reconstructions indeed may be performed as described, we develop a detailed computational model for such experiment and test our approach in-silico using actual wiring data for C. Elegans available from ssEM[8].

Using actual connectivity data from C. Elegans, we explicitly test how well circuit structure may be extracted with Brainbow + GRASP and how many trials may be necessary. We additionally consider a number of noise factors that may be present in real-life experiment. First, we consider the effect of possible biological variability in the circuit, i.e. variability in the count of synapses between the same neurons from one animal to another. We model this factor by modifying the connectivity matrix for each animal with a Poisson random component $C_{AB} \to C_{AB}(1-a_b) + P[a_b C_{AB}]$. Here, $P[x]$ stands for Poisson-distributed random number with mean x, and parameter a_b controls the degree of biological variability: a_b=0 corresponds to the case of no biological variability and a_b=1 corresponds to the case where synapses are formed completely at random with C_{AB} synapses on average between neurons *A* and *B*. Using connectivity matrix thus defined, we further assume that the samples of neurons *PRE* and *POST* are formed with Brainbow and each neuron is led to express pre- and post-synaptic split-GFP parts of GRASP with constant probability f=0.5. Subsequently, we assume that the count of GRASP puncta in such animal $Q(i)$ is obtained with light microscope. We allow a certain amount of error in each observation, which we model by Gaussian noise $Q(i) \to Q(i)(1 + a_o N[0,1])$. Here, $N[0,1]$ is Normally-distributed random number with zero mean and unit variance, and parameter a_o controls the amount of

noise in the measurements. Finally, the identities of neurons expressing GRASP in each trial are assumed to be determined with a computer algorithm. Possible error in cell-identification is modeled by shuffling all-to-all the identities of a small fraction a_i of neurons chosen at random. We note that this noise-model represents a worst-case scenario: in reality some cells would be better identified than the others and the confusions of identity would not be all-to-all. For this reason, we believe that a higher error-rate than that quoted in Figure 3B may be tolerated in actual experiments. A more detailed account of this noise factor may be in order as more information about statistics of this error becomes available in the future.

From such collection of in-silico Brainbow + GRASP trials, we reconstructed connectivity from different number of animals M from 500 to 12,000 by solving problem (3) with the method of alternate projections. We then asked how well reconstructed connectivity matrices corresponded to the original ssEM data. We inspected obtained reconstructions for different values of parameters a_b, a_o and a_i visually as a matrix (Supplementary Figure 1) and as scatter plots showing reconstructed vs. actual connection strengths (Supplementary Figure 2). Quantitatively, we characterized the reconstructions quality by the correlation coefficient r^2 between the reconstructed connection strengths and that in the original ssEM data. Figures 2B and D, and Figures 3A-D summarize the results of our experiments. We observe that full wiring diagram in C. Elegans may be indeed recovered under reasonable noise conditions from 5,000-10,000 measurements. We also observe that decreasing the fraction f of cells expressing GRASP in each animal may have an impact on the robustness of the reconstruction procedure to the observation noise (Figure 3D), although for the other

factors such as biological variability or cell misidentification such impact is insignificant (data not shown).

Thus, we conclude that the reconstruction of entire circuit in C. Elegans using Brainbow and GRASP may be attainable with already existing tools and technologies. Thanks to small size (100 μm diameter and 1000 μm long) and fast development (2 -3 days), 10,000 worms with given genotype may be rapidly incubated on a single Petri-dish. In fact, modern phenotype screens already routinely pay attention to populations of worms that large. The identities of the neurons expressing GRASP in each worm may be determined automatically from a high resolution optical scans of nuclei-targeted fluorescence with algorithm such as in Long et al.[25]. The measurement *Q* may be obtained by counting the number of distinct GRASP puncta from a high resolution scan. Or the combined fluorescence strength of GRASP puncta may be measured from a low resolution image of the worm, which may be a more robust measurement. Optical scans of C. Elegans may be performed in about 1 minute or less with off-the-shelf light microscopes (Kerr, R., personal communication). Given all of these, such reconstruction of full neural circuit in C. Elegans should be possible to complete in one week or less. The only substantial prerequisite for such an experiment is the creation of Brainbow + GRASP transgenic animal.

If, indeed, such reconstructions of complete neural circuit in C. Elegans may be performed as described, the ability to extract circuit routinely there already should prove valuable for systems neuroscience. For example, it would allow addressing many central questions about the relationship between circuit structure and behavior explicitly on the example of the nervous system of C. Elegans, as described in the introduction.

Described paradigm may be also applied in a variety of other circuits such as in Drosophila, zebrafish or mouse. Although here we specifically focused on Brainbow and GRASP, this paradigm may be adopted for use with different connectivity reporters such as direct tracing of neurons, ChR2 assisted circuit mapping or transsynaptic viruses, and either Brainbow, MARCM/MADM or UAS/Gal4 systems for driving expression patterns in Eq.(3b). Stochasticity of the expression patterns is not a required feature of this approach. On the contrary, nearly arbitrary set of expression patterns may be used to recover connectivity from Eqs.(3). We must also specifically point out that the ability to target individual or small groups of neurons, e.g. with UAS/Gal4, is not required. In particular, we showed that with broadly targeted connectivity probes one is able to effectively constrain the circuit in $O(N_p \log N)$ time. This should be compared with the minimal $O(N_p)$ time required to determine connectivity matrix from direct probes of connectivity between individual neurons, e.g. with whole-cell patches or other precisely targeted probes, when it is already a-priory known which pairs should be tested.

Although we showed that in C. Elegans such reconstruction may be performed with existing tools, applications in larger animals will require meeting a number of significant challenges. In larger animals unfavorable size scaling should be overcome: brains get bigger and so imaging gets harder, while the circuit complexity gets higher and so more measurements are needed. The focus on the development of transgenic models allowing for multiple connectivity probes from the same animal, such as multi-color GRASP, thus becomes highly interesting. Perhaps, even greater challenge is the necessity for independent identification of reporter's expression patterns in different animals. In larger animals it is

not known how to perform such identification or what the stereotypical units of neural connectivity may be, so the progress in the problems associated with charting neuronal populations, building brain atlases and the computational tools for identification of neurons and their classes gains added significance. Improvement of the reconstruction algorithm aimed at reduction of the minimal number of measurements or using different kinds of measurements may be also possible. While Candes et al.[26] provide a route for *exact* reconstruction of the connectivity matrix with theoretical performance bound, faster algorithms may be possible if certain amount of error is tolerated or if prior knowledge about the circuit structure may be accommodated. Development of such algorithms is an interesting direction for future theoretical research.

Acknowledgements

The author would like to acknowledge and is extremely grateful for the essential discussions with Shiv V. Vitaladevuni, Veit Elser, Max Nikitchenko and Liam Paninski. This work was partially carried out at Janelia Farm Research Campus, HHMI and Aspen Center for Physics.

Figure Captions

Figure 1: Circuit reconstruction using Brainbow + GRASP. A) Pre and post-synaptic split-GFP parts of GRASP (sGFP) are concatenated with two sequences coding for two nuclei-targeted fluorescent proteins (nFP), and both are flanked with inverted loxP sites. sGFP part of the construct allows producing large number of different GRASP connectivity probes from single genetic line, and nFP part of the construct allows identifying the cells thus

probed in each animal. B) In one Brainbow + GRASP animal, random subsets of neurons expressing pre-synaptic (red) and post-synaptic parts (green) of sGFP are selected with Cre from the above genetic construct. Whenever a synapse is present between any one from targeted pre- and post-synaptic neurons, sGFP may recombine and produce fluorescent puncta identifiable with light microscope (yellow). The total count of such puncta Q in each trial should be obtained. Such count mathematically corresponds to a constraint defining the sum of all entries at the intersections of selected pre-synaptic rows and post-synaptic columns in the connectivity matrix C (yellow). Three such trials are shown in the figure. When large number of trials is obtained, connectivity gets encoded in the fluctuations in Q in relation to GRASP expression patterns.

Figure 2: Results of circuit reconstruction from Brainbow + GRASP in C. Elegans, modeled in-silico from actual wiring diagram[8]. (A) Quality of the reconstructed connectivity matrix with triggered-average method, as measured by the correlation coefficient r^2 between reconstructed and actual connection strengths. (B) Quality of the reconstruction using the method of alternate projections. (C) Example of reconstruction from triggered-average method for M=300K measurements, r^2=0.4. (D) Example of reconstruction from the method of alternate projections for M=8K measurements, r^2=0.9.

Figure 3: Impact of different noise factors on circuit reconstruction, as measured by the correlation coefficient r^2 between reconstructed and actual connection strengths. (A) Biological variability is of least concern with very good reconstructions obtained even when synaptic strengths in each animal are nearly fully random. (B) Cell-identification errors have significant impact on reconstructions, implying that the fraction of misidentified

neurons in each trial should be kept below 3-5%. (C) Observation noise has significant impact on reconstructions, implying that the measurements should be obtained with relative error better than 3%. (D) Impact of observation noise may be effectively controlled by reducing GRASP expression frequency *f*, with already very large amounts of noise tolerated when *f*=0.1-0.2.

References

1. Broome, B. M., Jayaraman, V. & Laurent, G. Encoding and decoding of overlapping odor sequences. *Neuron* **51**, 467-482 (2006).
2. Jones, L. M., Fontanini, A., Sadacca, B. F., Miller, P. & Katz, D. B. Natural stimuli evoke dynamic sequences of states in cortical ensembles. *PNAS* **104**, 18772-7 (2007).
3. Rabinovich, M. I., Huerta, R., Varona, P. & Afraimovich, V. S. Transient cognitive dynamics, metastability, and decision making. *PLoS Comput. Biol.* **4**, e1000072 (2008).
4. Chalasani, S. H. et al. Dissecting a circuit for olfactory behavior in Caenorhabditis elegans. *Nature* **450**, 63-70 (2007).
5. Gray, J. M., Hill, J. J. & Bargmann, C. I. A circuit for navigation in Caenorhabditis elegans. *PNAS* **102**, 3184-91 (2005).
6. Dunn, N. A., Lockery, S. R., Pierce-Shimomura, J. T. & Conery, J. S. A neural network model of chemotaxis predicts functions of synaptic connections in the nematode Caenorhabditis elegans. *J. Comp. Neurosci.* **17**, 137-47 (2004).

7. Song, S., Sjostrom, P. J., Reiql, M., Nelson, S. & Chklovskii, D. B. Highly nonrandom features of synaptic connectivity in local cortical circuits. *PLoS Biol.* **3**, e68 (2005).

8. Wormatlas: a database of behavioral and structural anatomy of Caenorhabditis elegans at http://www.wormatlas.org.

9. Chen, B. L., Hall, D. H. & Chklovskii, D. B. Wiring optimization can relate neuronal structure and function. *PNAS* **103**, 4723-28 (2006).

10. Hall, D. H. & Russell, R. L. The posterior nervous system of the nematode Caenorhabditis elegans: Serial Reconstruction of identified neurons and complete pattern of synaptic interactions. *J. Neurosci.* **11**, 1-22 (1991).

11. White, J., Southgate, E., Thomson, J. N. & Brenner, S. The structure of the nervous system of the nematode caenorhabditis elegans. *Phil. Trans. Royal Soc. London. Series B, Biological Sciences* **314**, 1-340 (1986).

12. Jain, V. et al. Supervised learning of image restoration with convolutional networks. International Conference on Computer Vision, Rio de Janeiro, Brazil, 2007.

13. Jurrus, E. et al. Axon tracking in serial block-free scanning electron microscopy. MICCAI 2006 workshop., Copenhagen, 2006.

14. Macke, J. et al. Contour-propagation algorithms for semi-automated reconstruction of neural processes. *J. Neurosci. Methods* **167**, 349-57 (2008).

15. Mishchenko, Y. Automation of 3D reconstruction of neural tissue from large volume of conventional Serial Section Transmission Electron Micrographs. *J. Neurosci. Methods* In Press; doi:10.1016/j.jneumeth.2008.09.006 (2008).

16. Feinberg, E. H. et al. GFP Reconstitution Across Synaptic Partners (GRASP) defines cell contacts and synapses in living nervous systems. *Neuron* **57**, 353-63 (2008).

17. Lee, T. & Luo, L. Mosaic analysis with a repressible cell marker (MARCM) for Drosophila neural development. *Trends Neurosci.* **24**, 251-4 (2001).

18. Livet, J. et al. Transgenic strategies for combinatorial expression of fluorescent proteins in the nervous system. *Nature* **450**, 56-62 (2007).

19. Luo, L., Callaway, E. M. & Svoboda, K. Genetic dissection of neural circuits. *Neuron* **57**, 634-60 (2008).

20. Petreanu, L., Huber, D., Sobczyk, A. & Svoboda, K. Channelrhodopsin-2-assisted circuit mapping of long-range callosal projections. *Nat. Neurosci.* **10**, 663-8 (2007).

21. Phelps, C. B. & Brand, A. H. Ectopic gene expression in Drosophila using GAL4 system. *Methods* **4**, 367-79 (1998).

22. Wickersham, I. R., Finke, S., Conzelmann, K. K. & Callaway, E. M. Retrograde neuronal tracing with a deletion-mutant rabies virus. *Nat. Methods* **4**, 47-9 (2007).

23. Wu, J. S. & Luo, L. A protocol for mosaic analysis with a repressible cell marker (MARCM) in Drosophila. *Nat. Protocols* **1**, 2583-9 (2006).

24. Luo, L. Fly MARCM and mouse MADM: genetic methods of labeling and manipulating single neurons. *Brain Res. Rev.* **55**, 220-7 (2007).

25. Long, F., Peng, H., Liu, X., Kim, S. & Myers, E. Automatic recognition of cells (ARC) for 3D images of C. Elegans. in Lecture Notes in Computer Science: Research in Computational Molecular Biology pp. 128-139 (Springer Berlin, 2008).

26. Candes, E. J., Romberg, J. & Tao, T. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Inf. Theory* **52**, 489-509 (2006).

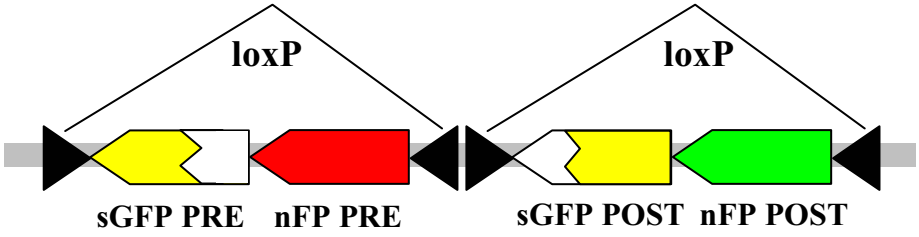
27. Vanderbei, R. J. *Linear Programming: Foundations and Extensions*. (Springer-Verlag, New York, 2001).

28. Candes, E. J. & Romberg, J. Practical signal recovery from random projections. SPIE Conference on Wavelet Applications in Signal and Image Processing XI, 2005.

29. Bregman, L. M. The method of successive projection for finding a common point of convex sets. *Soviet Math. Dokl.* **6**, 688-92 (1965).

Figure 1

A


loxP
loxP
sGFP PRE
nFP PRE
sGFP POST
nFP POST

B

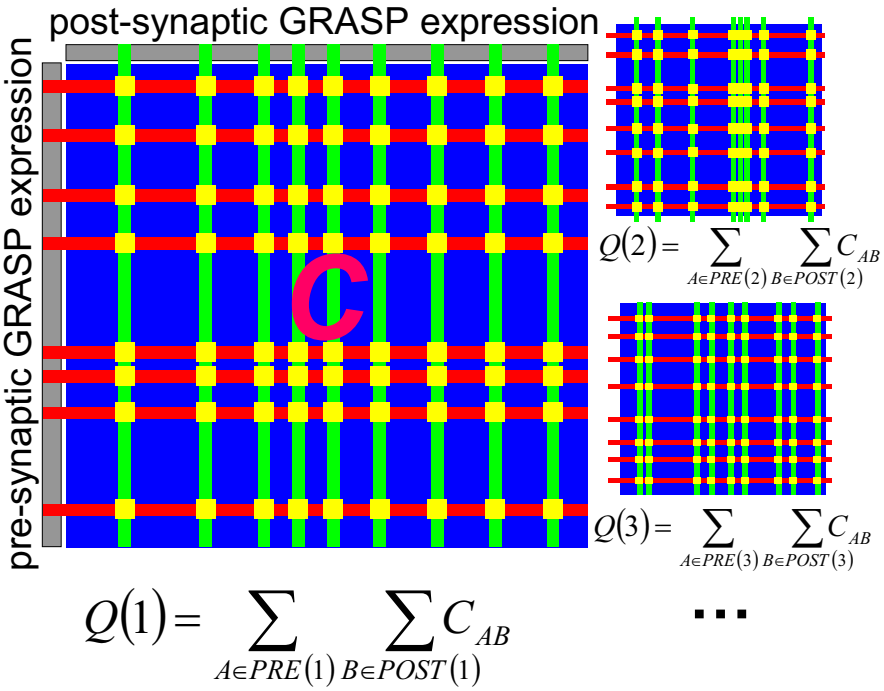

post-synaptic GRASP expression
pre-synaptic GRASP expression
C
$Q(1) = \sum_{A \in PRE(1)} \sum_{B \in POST(1)} C_{AB}$
$Q(2) = \sum_{A \in PRE(2)} \sum_{B \in POST(2)} C_{AB}$
$Q(3) = \sum_{A \in PRE(3)} \sum_{B \in POST(3)} C_{AB}$
...

Figure 2

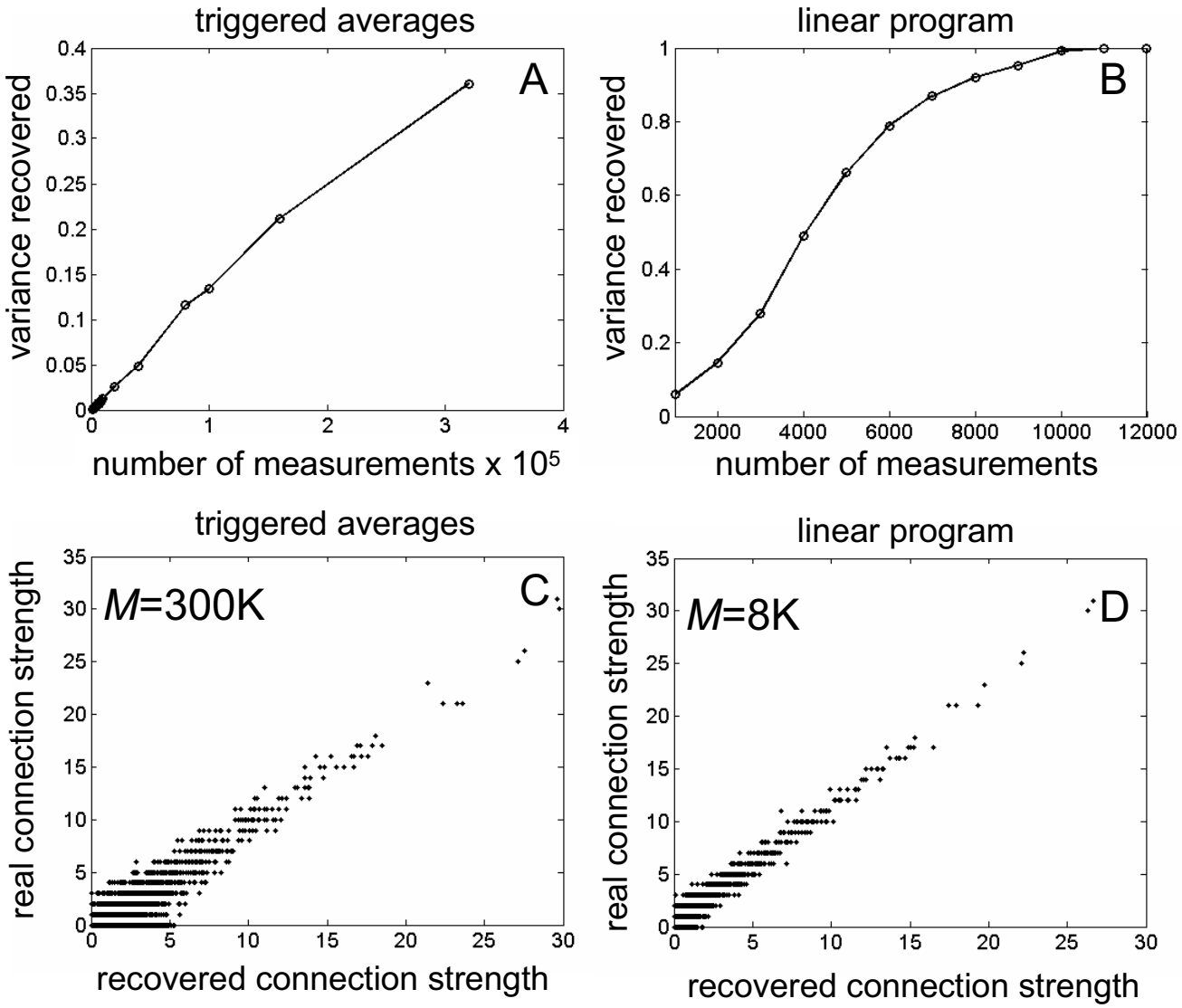
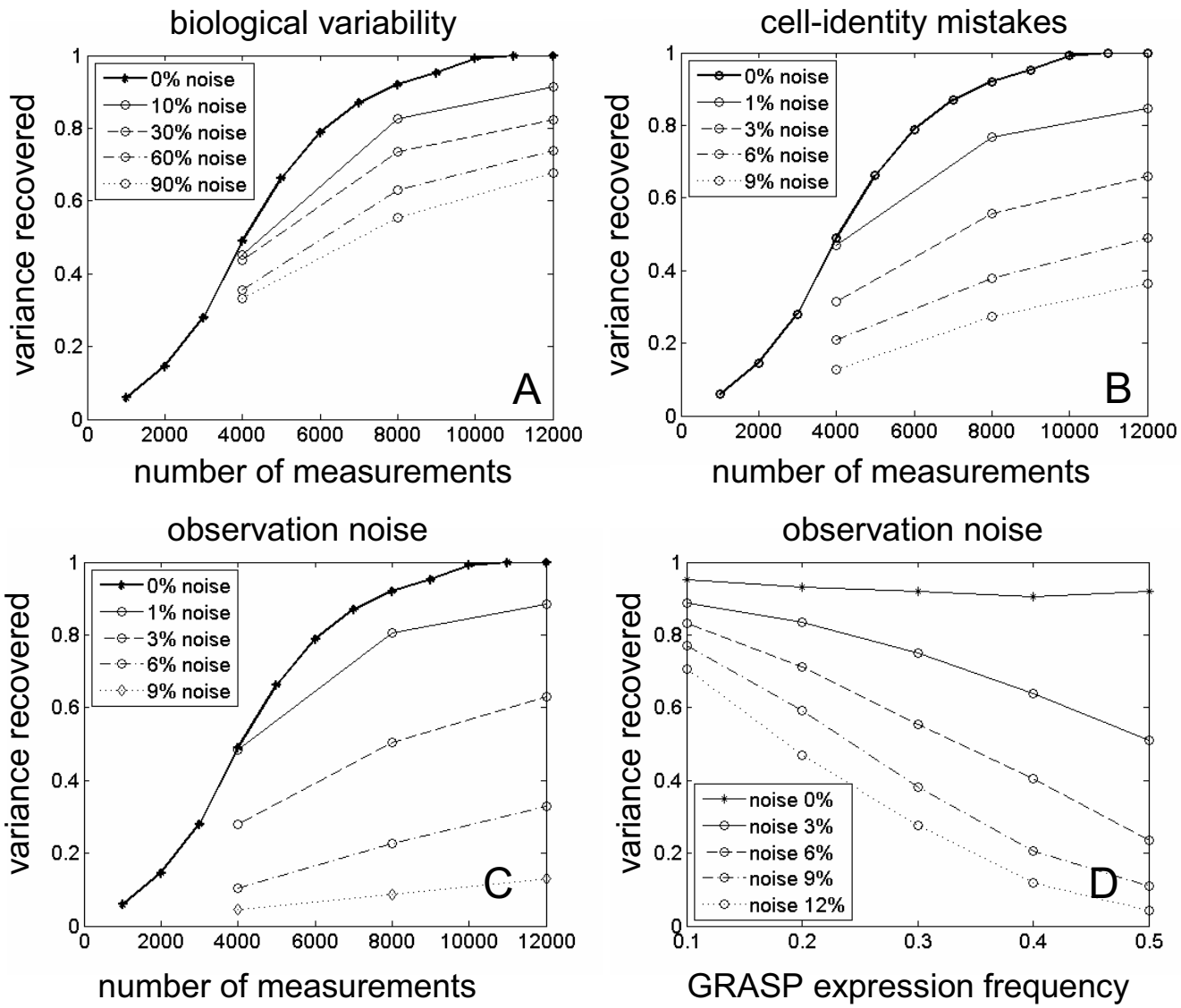

triggered averages
A
variance recovered
number of measurements x 10^5
linear program
B
variance recovered
number of measurements
triggered averages
M=300K
C
real connection strength
recovered connection strength
linear program
M=8K
D
real connection strength
recovered connection strength

Figure 3


biological variability
variance recovered
number of measurements
0% noise
10% noise
30% noise
60% noise
90% noise
A
cell-identity mistakes
variance recovered
number of measurements
0% noise
1% noise
3% noise
6% noise
9% noise
B
observation noise
variance recovered
number of measurements
0% noise
1% noise
3% noise
6% noise
9% noise
C
observation noise
variance recovered
GRASP expression frequency
noise 0%
noise 3%
noise 6%
noise 9%
noise 12%
D