Curriculum Vitae
Tao Hu

Howard Hughes Medical Institute	Tel: (571) 209-4000 ext. 3141 (Office)
Janelia Farm Research Campus	(612) 483-1334 (Cell)
19700 Helix Drive, Ashburn, VA 20147	Email: hut@janelia.hhmi.org

RESEARCH INTERESTS

1. Biophysically Motivated Modeling of Neural Sensory Coding
2. Computational Methods for Neural Circuit Reconstruction
3. Machine Learning Algorithms for Biological Data Mining

EDUCATION

2003-2008 School of Physics and Astronomy, University of Minnesota, Minneapolis, MN
Ph. D. in Physics
Dissertation title: "*How proteins search for their targets on DNA*"

1999-2003 School of Physics, Peking University, Beijing, China
B. S. in Physics

SCIENTIFIC EMPLOYMENT

09/2011-Present Research Specialist, Howard Hughes Medical Institute, Janelia Farm Research Campus

07/2008-08/2011 Postdoctoral Associate, Howard Hughes Medical Institute, Janelia Farm Research Campus

TEACHING EXPERIENCE

09/2003-05/2007 Teaching Assistant for the following courses at University of Minnesota:
Introductory Physics for Biology and Pre-medicine;
Introductory Physics for Science and Engineering

HONORS

1. Doctoral Dissertation Fellowship, 2007-2008, Graduate School, Univ. of Minnesota.
2. Anatoly Larkin Fellowship, 2007, School of Physics & Astronomy, Univ. of Minnesota.

PUBLICATION LIST

Papers in Preparation:

1. **Tao Hu** and D.B. Chklovskii, “Predictive coding of time-varying sensory stimuli”.

Papers Submitted:

1. S. Druckman***, Tao Hu* (* equal contribution)** and D.B. Chklovskii, “Non-linear predictive coding as a model of early sensory processing”.
2. **Tao Hu**, J. Nunez-Iglesias, S. Vitaladevuni, L. Scheffer, Shan Xu, M. Bolorizadeh, H. Hess, R. Fetter and D.B. Chklovskii, “Super-resolution reconstruction of brain structure using sparse representation over learned dictionary”.
3. **Tao Hu** and D.B. Chklovskii, “Sparse LMS via online linearized Bregman iteration”.

Papers Published:

1. **Tao Hu**, Alexander Genkin and D.B. Chklovskii, “Computing sparse representations using a network of sparsely communicating nodes”, accepted to Neural Computation, (2012).
2. D. Glasner, **Tao Hu**, J. Nunez-Iglesias, L. Scheffer, Shan Xu, H.F. Hess, R. Fetter, D.B. Chklovskii and R. Basri, “High resolution segmentation of neuronal tissues from low depth-resolution EM imagery”, Lecture Notes in Computer Science 6819 , 261 (2011).
3. Y. Mishchenko*, **Tao Hu* (* equal contribution)**, J. Spacek, J. Mendenhall, K.M. Harris and D.B. Chklovskii, “Ultrastructural analysis of hippocampal neuropil from the connectomics perspective”, Neuron 67, 1009, (2010).
4. **Tao Hu** and D.B. Chklovskii, “Reconstruction of sparse circuits using multi-neuronal excitation (RESCUME)”, Advances in Neural Information Processing Systems (NIPS) 22, 790 (2009).
5. **Tao Hu** and B.I. Shklovskii，“Theory of DNA translocation through narrow ion channels and nanopores with charged walls”, Physical Review E 78, 032901 (2008).
6. **Tao Hu**, R. Zhang and B.I. Shklovskii，“Electrostatic theory of viral self-assembly: a toy model”, Physica A 387, 3059 (2008).
7. **Tao Hu** and B.I. Shklovskii，“How a protein searches for its specific site on DNA: role of intersegment transfer”, Physical Review E 76, 051909 (2007).
8. **Tao Hu** and B.I. Shklovskii，“Kinetics of viral self-assembly: the role of ss RNA antenna”, Physical Review E 75, 051901 (2007).

9. **Tao Hu** and B.I. Shklovskii，“Hopping conductivity of a suspension of flexible wires in an insulator”, Physical Review B 74, 174201 (2006).

10. **Tao Hu** and B.I. Shklovskii，“Theory of hopping conductivity of a suspension of nanowires in an insulator”, Physical Review B 74, 054205 (2006).

11. **Tao Hu** and B.I. Shklovskii，“How does a protein search for the specific site on DNA: the role of disorder”, Physical Review E 74, 021903 (2006).

12. **Tao Hu**, A.Yu. Grosberg and B.I. Shklovskii，“Conductivity of a suspension of nanowires in a weakly conducting medium”, Physical Review B 73, 155434 (2006).

13. **Tao Hu**, A.Yu. Grosberg and B.I. Shklovskii，“How proteins search for their specific sites on DNA: the role of DNA conformation”, Biophysical Journal. 90, 2731 (2006).

14. Y.J. Sun, W.S. Gao，**Tao Hu** and Q.J. Xing, “Photoelastic Waveguides Induced by Thin Film Composite Structure in InGaAsP/InP Double Heterostructures”, Semiconductor Science and Technology 21, 575 (2006).

15. **Tao Hu**, X. Lu, Y. Yan, Y. Fu, H. Zhang and S. L. Zhang, “Raman Spectroscopic Study on Iron Oxides Prepared by Oxidation of Iron”, Spectroscopy and Spectral Analysis 24, 1072 (2004).

16. **Tao Hu**, Y. Fu, H. Zhang, S.L. Zhang, J. Ouyang and X.S. Zhao“Temperature-dependent Raman Spectra of α-Fe_2O_3 Nano-wire and Bulk Material”, Chinese Journal of Light Scattering 16, 303 (2004).

17. W.S. Gao, Q.J. Xing, Z.J. Yuan and **Tao Hu**, “W0.95Ni0.05 Compressive Strain Thin Film Stripe Window Induced Photoelastic Waveguide Structure in InGaAsP/InP Double Heterostructure”, Chinese Physics Letters 20, 1296 (2003).

Book:

Tao Hu, “Visual C++.NET Programming Experience”, Electronic Industry Press, Beijing, China 2003. (In Chinese)

CONFERENCE ABSTRACTS/PRESENTATIONS

1. “Predictive coding of time-varying sensory stimuli”, *Neuronal Circuits meeting*, CSHL, NY, Mar. 2012.

2. “Biologically plausible learning of sparse-coding dictionary in a neural network”, *Computational and Systems Neuroscience (Cosyne),* Salt Lake City, Utah, Feb. 2012.

3. "Computing sparse representations using a network of integrate-and-fire neurons", *Computational and Systems Neuroscience (Cosyne),* Salt Lake City, Utah, Feb. 2012.

4. "Reconstruction of brain circuits using computational super-resolution", *High Resolution Circuit Reconstruction*, Janelia Farm Research Campus, Howard Hughes Medical Institute, Ashburn, Virginia, Sep., 2011.

5. "Computing sparse representations using a network of integrate-and-fire neurons", *Sloan-Swartz 2011 Annual Meeting,* Janelia Farm Research Campus, Howard Hughes Medical Institute, Ashburn, Virginia, Feb., 2011.

6. "Early sensory processing as predictive coding: subtracting sparse approximations by circuit dynamics", *Computational and Systems Neuroscience (Cosyne),* Salt Lake City, Utah, Feb. 2011.

7. "Super-resolution reconstruction of brain structure", *Sampling and Reconstruction: Applications and Advances*, Banff International Research Station for Mathematical Innovation and Discovery (BIRS), Banff, Alberta, Canada, Nov., 2010.

8. "Estimating conectomes with compressed sensing", *Emerging Techniques in Neuroscience*, The Kavli Institute for Theoretical Physics, UCSB, Oct. 2010.

9. "Reconstruction of sparse circuits using multi-neuronal excitation (RESCUME)", *Computational and Systems Neuroscience (Cosyne),* Salt Lake City, Utah, Feb. 2010.

10. "Reconstruction of sparse circuits using multi-neuronal excitation (RESCUME)", *The Neural Information Processing Systems (NIPS),* Vancouver, BC, Canada, Dec. 2009.

11. **"**Kinetics of viral self-assembly: the role of ss RNA antenna"**,** *Physics Inspired by Biology,* The Fine Theoretical Physics Institute, University of Minnesota, May., 2007.

12. "Electrostatic theory of viral self-assembly: Structure and Kinetics"**,** *The American Physical Society (APS) March Meeting*, Denver, Colorado, Mar., 2007.

REFERENCES

- Dmitri Chklovskii, Group Leader
  Janelia Farm Research Campus, HHMI
  19700 Helix Drive, Ashburn, VA 20147
  Phone: 571-209-4000
  Email: chklovskiid@janelia.hhmi.org

- Boris Shklovskii, A.S. Fine Chair in Theoretical Physics
  School of Physics and Astronomy, University of Minnesota
  116 Church Street SE, Minneapolis, MN 55455

Phone: 612-625-0771
Email: shklovsk@physics.umn.edu

- Henry Greenside, Professor in Physics
  Department of Physics, Duke University
  Physics Bldg., Science Dr. Durham, NC 27708
  Phone: 919-660-2548
  Email: hsg@phy.duke.edu

RESEARCH STATEMENT

The central problem that I am pursuing is how neural circuits lead to function. By relying on recent advances in statistics, signal processing and machine learning techniques, I focus my research on developing computational tools for neural circuit reconstruction (the connectomics) [1-4] and biophysically motivated models for neural information processing [5, 6]. I applied compressive sensing to improve the efficiency of electrophysiological approach for probing sparsely connected neural circuit [1]. I also developed an unsupervised learning algorithm to improve the depth resolution of electron microscopy, which greatly improves the ability of tracing neural processes automatically [3, 4]. I believe for the understanding of brain function, not only reconstructing brain wiring diagrams is necessary, but constructing theoretical models to infer the computation from neural circuitry is also indispensable. In this direction, I developed a biophysically plausible spiking neural network for sparse coding [5], and a non-linear predictive coding model for information processing in early sensory systems such as retina and olfactory bulb in vertebrates or antennal lobes in invertebrates [6]. These topics provide ample opportunities to participate in collaborative, interdisciplinary research addressing the fundamental issues of neural information processing. Below is a summary of my recent research and the agenda in the future.

CURRENT RESEARCH

1. Biophysical Models of Sensory Coding

1.1 Early sensory processing as non-linear predictive coding

Early stages of sensory systems face the challenge of compressing information from a large number of receptors onto a much smaller number of projection neurons, a so-called communication bottleneck. To make more efficient use of limited bandwidth, compression may be achieved using predictive coding, whereby predictable or redundant components of the stimulus are removed. In the case of the retina, Srinivasan et al. [7] suggested that feed-forward subtraction of a linear prediction generated from nearby receptors implements such compression, resulting in biphasic center-surround receptive fields. However, inhibition often operates in a feedback manner and with non-linear input output transformations, considerably complicating the dynamics of such circuits. I solved the transient *non-linear* recurrent dynamics of a generic early sensory circuit in response to a step-like stimulus [6]. I showed that interneuron activity in time constructs progressively less sparse but more accurate representations of the stimulus, thus providing a powerful theoretical framework to understand the dynamics of early sensory processing in a variety of physiological experiments. I proved that threshold-linear neurons are superior to both linear neurons and direct transmission. More generally, our results demonstrate that highly non-trivial computations, at the forefront of modern signal processing, can be mapped onto a concrete neuronal circuit.

1.2 Sparse coding using a spiking neural network

Many natural signals can be sparsely represented as linear combinations of a few feature vectors (or elements) chosen from an over-complete dictionary. The importance of sparse representations has long been recognized in applied mathematics and in neuroscience. In applied mathematics, sparse representations lie at the heart of many important developments. In signal processing, such solutions serve as a foundation for basis pursuit de-noising, compressive sensing and object recognition. In statistics, regularized multivariate regression algorithms, such as the Lasso or the elastic net, rely on sparse representations to perform feature subset selection along with coefficient fitting. In neuroscience, electrophysiological recordings and theoretical arguments demonstrate that most neurons are silent at any given moment. Given the ubiquity of sparse coding, how neural networks compute sparse representations remains a challenging question. I proposed a hybrid distributed algorithm (HDA), which computes sparse redundant representations using a biologically inspired network of integrate-and-fire neurons [5]. HDA neurons perform both gradient-descent-like steps on analog internal variables, i.e. the sub-threshold membrane potentials, and coordinate-descent-like steps via quantized external variables communicated to each other, i.e. the spike trains. I proved the convergence of HDA and showed that it is stable against time-varying noise, specifically, the representation error decays as $\sqrt{t}$ for Poisson noise. Since HDA works on a network of simple nodes communicating via low-bandwidth channels, it may also be useful for sensor networks with limited energy budget.

2. Computational Methods for Neural Circuit Reconstruction and Statistical Analysis on Brain Tissue Ultrastructures

2.1 Reconstructing sparse circuits using multi-neuronal excitation (RESCUME)

Synapses onto a neuron can be probed by sequentially stimulating potentially pre-synaptic neurons while monitoring the membrane potential of the post-synaptic neuron. Reconstructing a large neural circuit using such a “brute force” approach is rather time-consuming and inefficient because the connectivity in neural circuits is sparse. Instead, I proposed an approach [1] which measures a post-synaptic neuron’s membrane potential while stimulating sequentially random subsets of multiple potentially pre-synaptic neurons. To reconstruct these synaptic connections from the recorded potential I applied a decoding algorithm recently developed for compressive sensing. Compared to the brute force approach, our method promises significant time savings that grow with the size of the circuit. Multi-neuronal stimulation allows reconstructing synaptic connectivity just from the spiking activity of post-synaptic neurons, even when sub-threshold potential is unavailable. I also applied our method to map the receptive fields of retinal ganglion cells from recorded sub-threshold potentials or spike trains [1]. Compared to the conventional reverse correlation or spike trigged average approaches, our method requires much less recording time, which is valuable for studying some fast adaptation problems.

2.2 Ultrastructural analysis of hippocampal neuropil from the connectomics perspective

Smaller invertebrate circuits can be reconstructed using serial section transmission electron microscopy (ssTEM) by identifying synapses and manually tracing pre- and postsynaptic neuronal processes to their cell bodies as has been demonstrated for the C. elegans nervous system. However, reconstructing vertebrate circuits using ssTEM manually is impractical. I automated registration and segmentation for ssTEM and fully reconstructed an unprecedented volume of 670 μm^3 from the rat hippocampus [2]. Although the reconstructed volumes are too small to contain complete circuits, they demonstrate that ssTEM can be scaled through automation. In addition, I used the reconstructed volumes as proving grounds to determine whether other approaches based on proximities between axons and dendrites can yield reliable predictions of synaptic connectivity. I found, first, in contrast to Peters' rule, the density of axons within reach of dendritic spines did not predict synaptic density along dendrites because the fraction of axons making synapses was variable. Second, an axo-dendritic touch did not predict a synapse; nevertheless, the density of synapses along a hippocampal dendrite appeared to be a universal fraction, 0.2, of the density of touches. Finally, the largest touch between an axonal bouton and spine indicated the site of actual synapses with about 80% precision but would miss about half of all synapses.

2.3 Limited angle tomography: neural circuit reconstruction using computational super-resolution

Reconstructing neuronal circuits on the synaptic level is a central problem in neuroscience. Large range of scales in brain architecture requires both high-resolution and high-throughput imaging. Existing electron microscopy (EM) techniques possess required resolution in the lateral plane and either high-throughput or high depth-resolution but not both. I exploited recent advances in unsupervised learning and signal processing to obtain high depth-resolution EM images computationally without sacrificing throughput. I found that the brain tissue can be represented as a sparse linear combination of localized basis functions that are learned using high-resolution datasets. Inspired by compressive sensing, I developed a technique called limited angle tomography that can reconstruct the brain tissue from very few tomographic views (typically 5) of each section [3]. In addition, I developed a method for detecting the high resolution locations of membranes directly from low depth-resolution images by learning a discriminative, over-complete dictionary [4]. These works facilitate tracing of neuronal processes and, hence, high throughput reconstruction of neural circuits to the level of individual synapses.

3. Signal Processing Algorithm

Better understanding of neural information processing also helps developing novel signal processing algorithms. For example, the above-mentioned HDA both serves as a model for neural computation and a method implementable to practical sensor networks. Inspired by sparse coding and online-learning, I proposed a sparse version of the least-mean-square (LMS) algorithm for adaptive signal processing called online linearized Bregman iteration (OLBI) [8]. Many signal processing tasks, such as signal prediction, noise cancellation or system identification, can be solved using adaptive linear filters. LMS is one of the most popular algorithms for adjusting filter weights. I considered signals whose true filter is sparse, containing only a few non-zero weights. I derived OLBI in the online learning framework by using the Follow-The-Regularized-Leader strategy and considering a non-differentiable sparsity inducing regularizer, the l_1-l_2 norm known as elastic net. I demonstrated numerically that the performance of OLBI is superior to that of the recently developed sparse LMS algorithms.

FUTURE RESEARCH PLANS

Through existing and new collaborations, I will move forward in developing a long-term interdisciplinary research program that contributes to the understanding of fundamental issues of neural information processing. I will continue to explore the principles underlying early sensory information coding and would like to expand my research to the hierarchical processing of progressively specialized and invariant features by cortical neurons. I am also interested in the computational issues that arise in the study of motor control and motor learning, e.g. the functional roles of neuronal redundancy which results in an over-complete representation of muscle activities. These topics are synergistic with the other focus of my interest, the connectomics, since the computational models both rely on the information from wiring diagrams and sharpen the questions addressed to them. In this direction, I want to develop computational methods for neural imaging and data analysis. Here I describe a few directions on which I expect to work in the near future.

1. Sensory Coding of Time-varying Stimuli

Natural time-varying stimuli are highly redundant, possessing significant temporal correlations over a wide range of time scales. For example, a turbulent odor plume contains a scale-free distribution of fluctuations in odor concentration. How neurons encode such stimuli efficiently remains a challenging problem for sensory neuroscience. In our recent work [6], I considered how a population of non-linear graded neurons encodes a step stimulus. I would like to extend the theoretical framework of predictive coding [6] and sparse coding [5] to investigate the optimal coding of naturalistic signal varying on multiple time scales, and develop a spiking neural network model which captures the key features of sensory processing in early sensory systems, such as retina and olfactory bulb in vertebrates or antennal lobes in invertebrates.

2. Computational Super-resolution for Light Microscopy

I plan to extend my research on computational super-resolution for electron microcopy [3] to light microcopy. Light microscopy is ideal for imaging live samples. However, its resolution is limited by diffraction; objects which are closer than half the wavelength of the light are not distinguishable. In general, the high spatial frequency in the sample lost in the microscope is unrecoverable. However, in many practical applications, biological samples are sparsely labeled fluorescently. Inspired by compressive sensing, I intend to develop sparsity regularized de-convolution techniques to improve the resolution of conventional light microscope. Such technique may also be valuable for improving the throughput of physical super-resolution techniques, e.g. photoactivated localization microscopy (PALM). If multiple molecules can be reliably localized in one point spread function, one can take fewer images each with more molecules to get a PALM image containing the same number of localized molecules.

3. Online Sparse Logistic Regression

Biomedical data, for example the microarray data and fMRI data, usually comprise large number of samples with highly correlated features. Under such condition, many common un-regularized batch learning approaches fail because of the high demands on memory storage and computation efficiency, and the potential danger of over-fitting. A widely used data mining technique is logistic regression. I am interested in developing an online logistic regression algorithm, incorporating the ℓ_1-regularization to promote sparse feature selection and thus make the classifier and classification decisions more comprehensible to people.

REFERENCES

1. **Tao Hu** and D.B. Chklovskii, “Reconstruction of sparse circuits using multi-neuronal excitation (RESCUME)”, Advances in Neural Information Processing Systems (NIPS) 22, 790 (2009); additional tests on actual experimental data presented as a poster in Cosyne 2010.

2. Y. Mishchenko*, **Tao Hu* (* equal contribution)**, J. Spacek, J. Mendenhall, K.M. Harris and D.B. Chklovskii, "Ultrastructural analysis of hippocampal neuropil from the connectomics perspective", Neuron 67, 1009, (2010).
3. **Tao Hu**, J. Nunez-Iglesias, S. Vitaladevuni, L. Scheffer, Shan Xu, M. Bolorizadeh, H. Hess, R. Fetter and D.B. Chklovskii, "Super-resolution reconstruction of brain structure using sparse representation over learned dictionary", submitted.
4. D. Glasner, **Tao Hu**, J. Nunez-Iglesias, L. Scheffer, Shan Xu, H.F. Hess, R. Fetter, D.B. Chklovskii and R. Basri, "High resolution segmentation of neuronal tissues from low depth-resolution EM imagery", Lecture Notes in Computer Science 6819 , 261 (2011).
5. **Tao Hu**, A. Genkin and D.B. Chklovskii, "Computing sparse representations using a network of integrate-and-fire neurons", accepted to Neural Computation, (2012).
6. S. Druckman***, Tao Hu* (* equal contribution)** and D.B. Chklovskii, "Non-linear predictive coding as a model of early sensory processing", submitted; also presented as a poster in Cosyne 2011.
7. M.V. Srinivasan, S.B. Laughlin and A. Dubs, "Predictive coding: a fresh view of inhibition in the retina", Proc R Soc Lond, B, 216, 427, (1982).
8. **Tao Hu** and D.B. Chklovskii, "Sparse LMS via online linearized Bregman iteration", submitted, (2012).

TEACHING STATEMENT

Becoming a professor at a university has always been my career goal, not only because I love the excitement of doing research, but also because I enjoy the pleasure and challenge of teaching. While I was a Ph.D. student at the University of Minnesota, I was lucky enough to have the opportunity to teach several introductory physics courses as a teaching assistant during academic years 2003-2007. These four years of teaching experience have affirmed my passion for teaching and significantly deepened my understanding of teaching.

At the core of my teaching philosophy lies my concern for students. In order to better understand their needs and feelings, I always try to recall my own experience as a student studying the same material. At the beginning of each course, I collect information from students to know more about their backgrounds and their purpose for taking the course. Besides the course evaluation required by the university usually at the end of the semester, I also administer anonymous course evaluations in the middle of the course so that I can recognize any problems I have and adjust my teaching methods in time. Depending on students' needs and backgrounds, the subject matter of the course and the size of the class, my teaching methods vary. However, no matter what course I teach, I always keep the following three principles in mind.

First of all, I believe interactions with students are very important. According to cognitive neuroscience, meaningful engagement with information is far more effective in the formation of long-term memory than mere passive listening. Therefore, I endeavor to encourage my students to actively participate in class. For example, I usually initiate a class with questions inviting students to think by themselves before I lecture. Whenever possible, I engage students in group work, discussion or projects. I also encourage them to interrupt me at any time with any comments or questions. I welcome them to challenge me or other authorities, which usually elicits stimulating discussions and also gives me inspiration for my own research from time to time.

Moreover, as a teacher, my role is not simply to impart factual knowledge; more importantly, my job is to provide the means for students to acquire knowledge independently and to utilize learned ideas and methods in new situations. I want my students to view the learning of science as a process of problem-solving and discovery rather than the ingestion of facts. Therefore, I strive to stimulate students' logical thinking, critical thinking, problem-solving, collaboration and communication skills, which will be imperative for their success no matter what field they ultimately choose. When students ask me how to solve a problem, I usually don't tell them the answer directly. Rather, I teach them how to break a bigger problem into several smaller steps and provide them with hints of possible methods and necessary sources to refer to. Usually, students will come back with a much deeper understanding of the subject and a sense of achievement after they accomplish the task by themselves.

Last but not least, I devote myself to fostering students' interest in the subject of the course. There is nothing better than the feeling of transferring my passion for science to my students. I feel that my students can sense my enthusiasm and become more interested in the subject themselves. To keep the course interesting and relevant for both majors and non-majors, I also relate the course content to real world examples and present related illustrations and videos whenever applicable. When there are students who show interest in specific topics, I always encourage them to do further study or research, and I am willing to offer guidance and assistance on additional readings and projects. I hope students can recognize the beauty and elegance of science as I do and even become scientists themselves one day.