Biological Blueprints for Human Inspired AI

Stanford, April 7 & 9, 2020

Acknowledgements: Nikhil Bhattasali, Chaofei Fan, Eugene Lewis, Megumi Sano, Yash Savani and Rishabh Singh

Biological Blueprints for Human Inspired AI[‡]

What does it mean for a brain to perform computations?[†]

[†] The first of several naive rhetorical questions we'll ask in this lecture.

[‡] Part I: Stanford CS379C © 2020 Thomas Dean

There are Many Approaches to Studying the Brain

Blind monks examining an elephant, an Ukiyo-e print by Hanabusa Itchō (1652–1724).

Adam Marblestone

Brenden Lake

Jessica Hamrick

Jill Leutgeb

Lisa Giocomo

Loren Frank

Matt Botvinick

Michael Frank

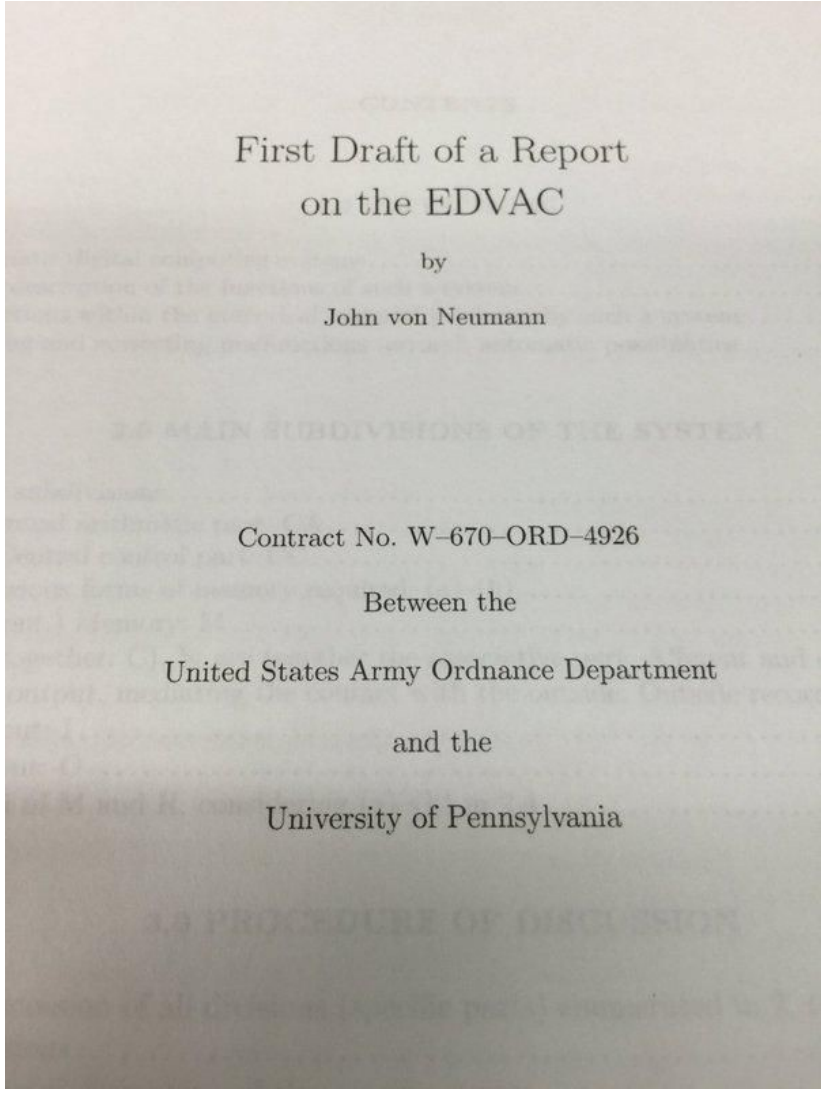
Oriol Vinyals

Randy O'Reilly

Peter Battaglia

Vivek Jayaraman

Birth of the Modern Computer: The von Neumann Architecture



First Draft of a Report
on the EDVAC

by

John von Neumann

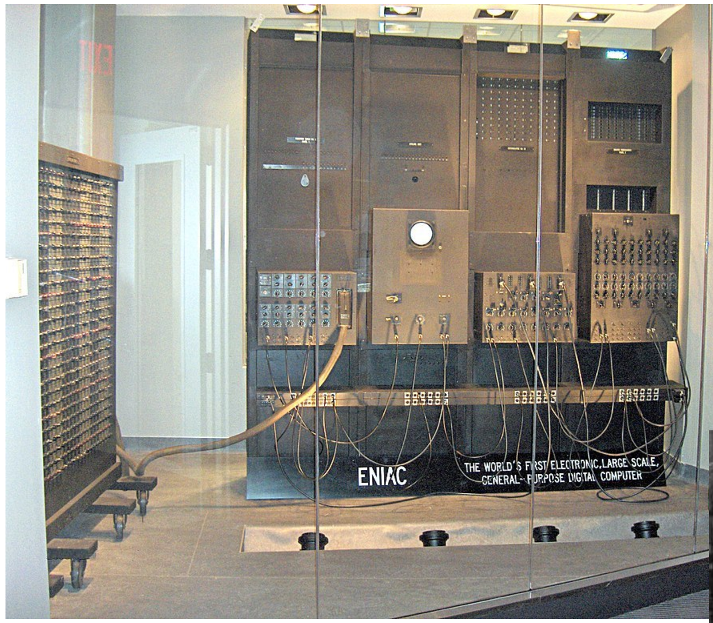
Contract No. W–670–ORD–4926

Between the

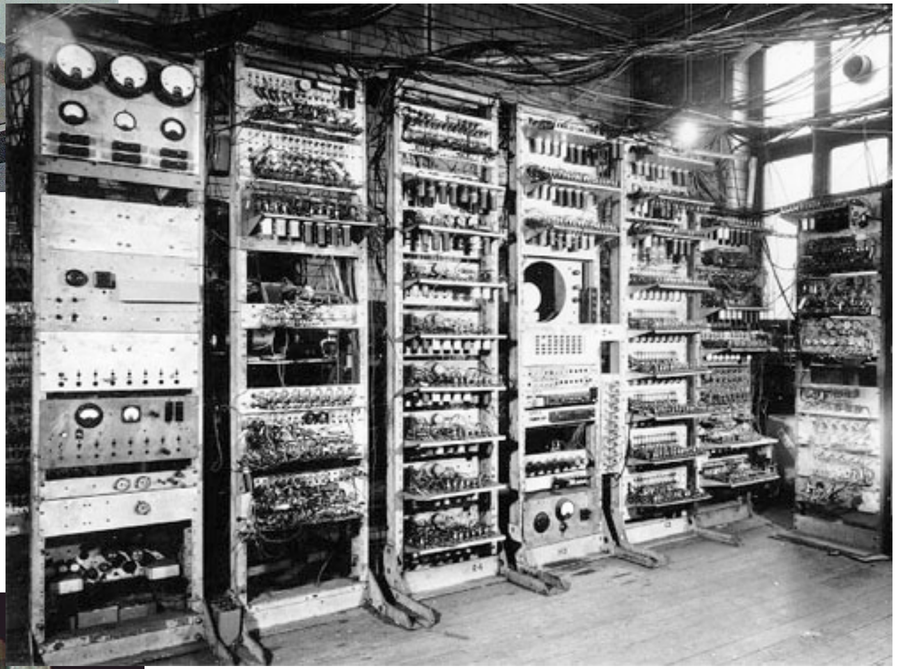
United States Army Ordnance Department

and the

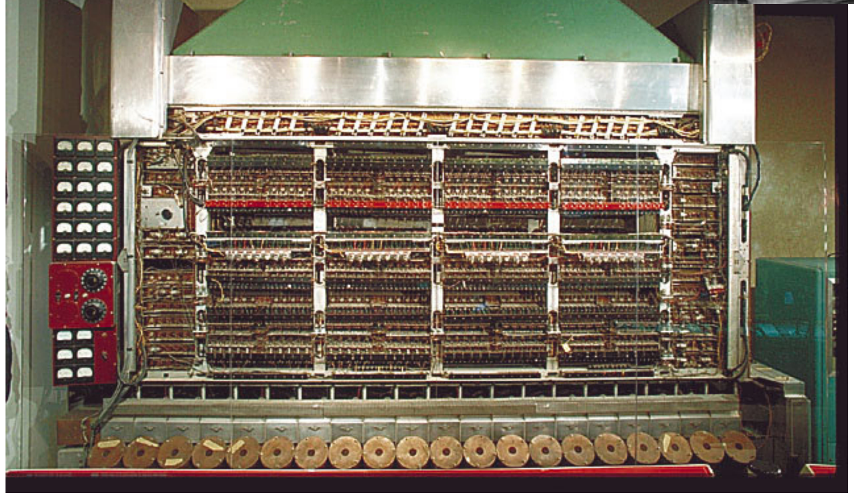
University of Pennsylvania


ENIAC
THE WORLD'S FIRST ELECTRONIC, LARGE-SCALE, GENERAL-PURPOSE DIGITAL COMPUTER

ENIAC



EDVAC

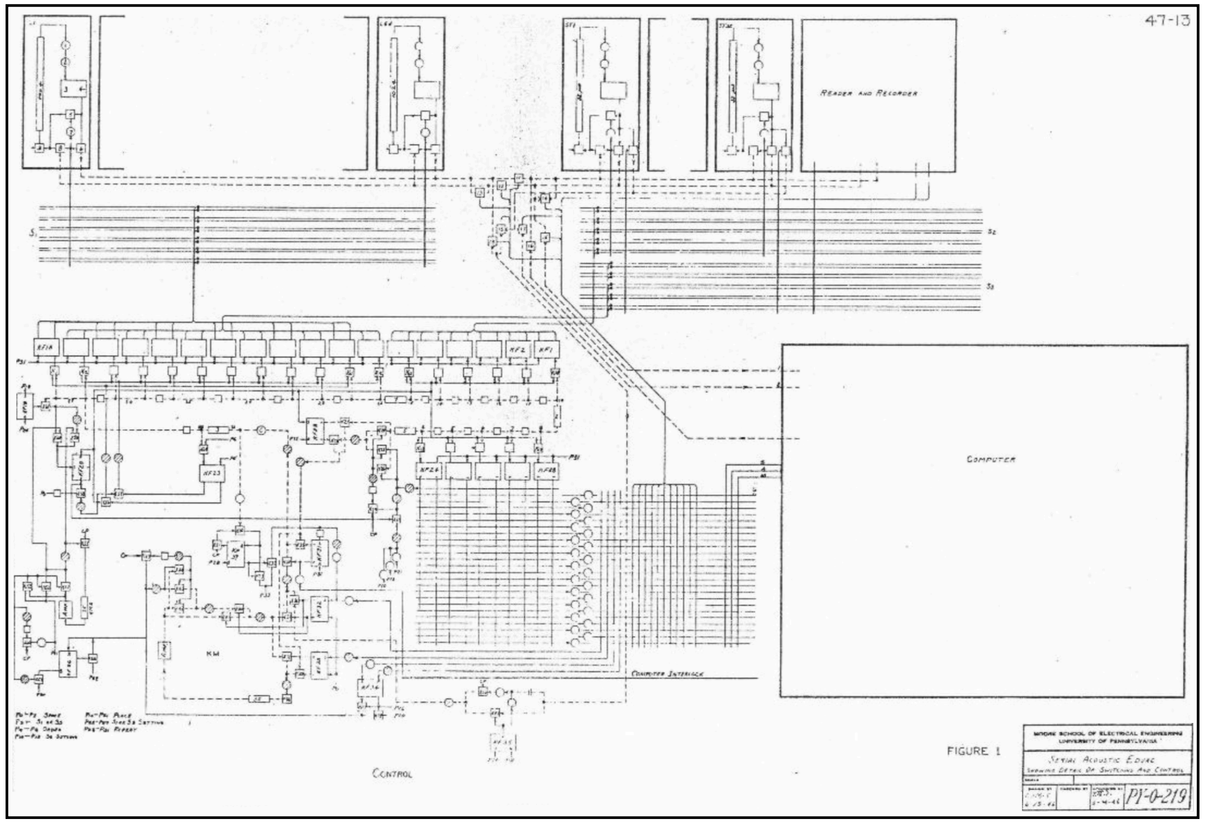


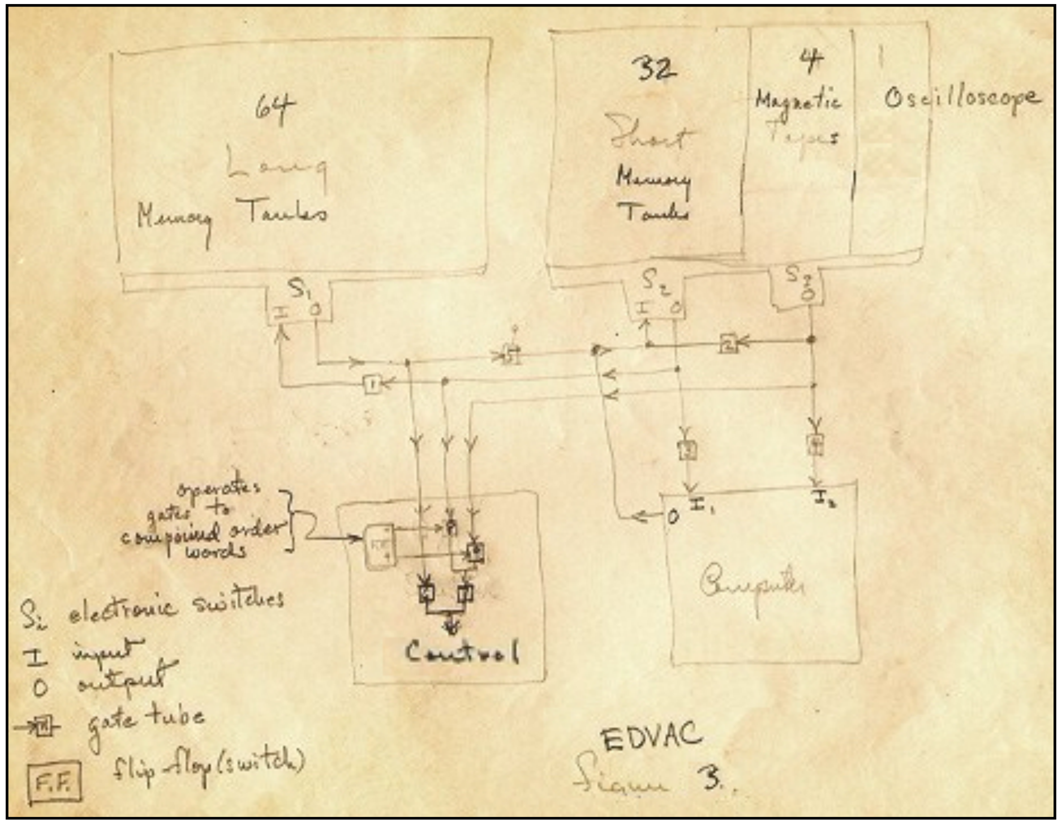
IAS

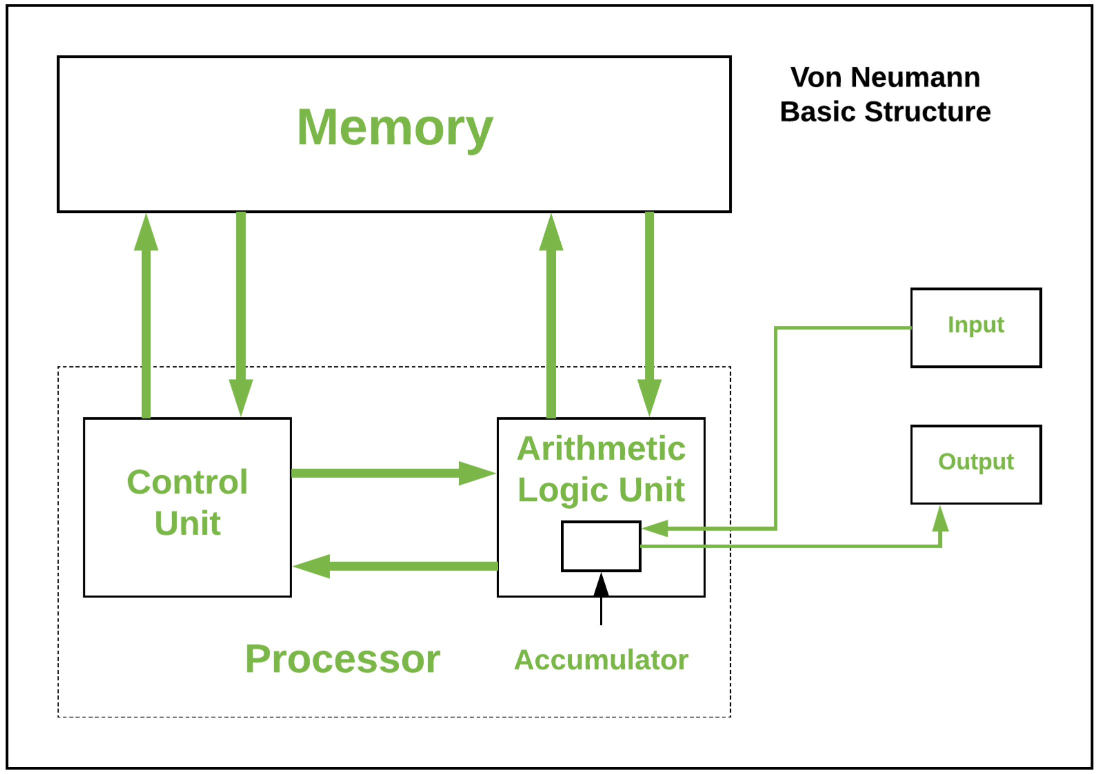
Alice Wang, Benton Calhoun, and Anantha Chandrakasan. *Sub-Threshold Design for Ultra Low-Power Systems.* Springer-Verlag, 2006.

What would a human-brain analog of von Neumann’s Architecture look like?

Birth of the Modern Computer: The von Neumann Architecture


47-13
READER AND RECORDER
COMPUTER
CONTROL
FIGURE 1
MOORE SCHOOL OF ELECTRICAL ENGINEERING
UNIVERSITY OF PENNSYLVANIA
SERIAL ACOUSTIC EDVAC
SHOWING DETAIL OF SWITCHING AND CONTROL
PY-0-219





Von Neumann Basic Structure

Memory

Control Unit

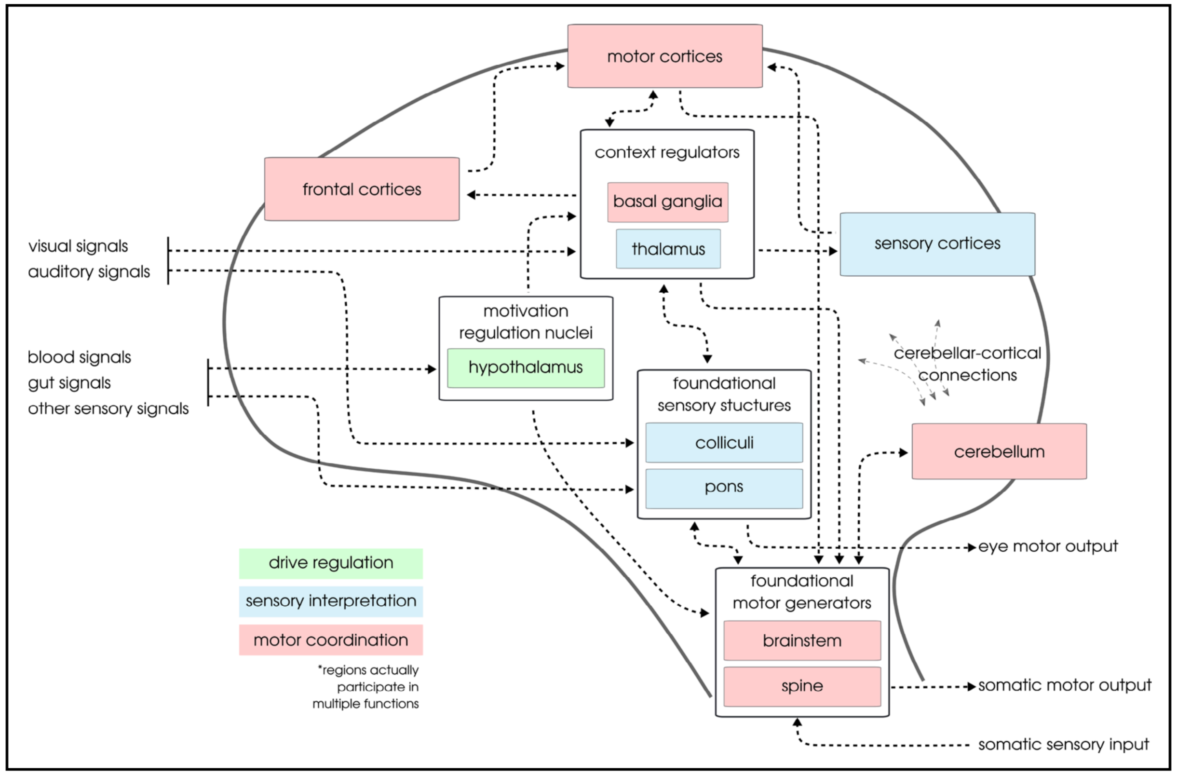
Arithmetic Logic Unit

Input

Output

Processor

Accumulator



motor cortices

frontal cortices

context regulators

basal ganglia

thalamus

sensory cortices

visual signals

auditory signals

motivation regulation nuclei

hypothalamus

blood signals

gut signals

other sensory signals

foundational sensory stuctures

colliculi

pons

cerebellar-cortical connections

cerebellum

eye motor output

drive regulation

sensory interpretation

motor coordination

*regions actually participate in multiple functions

foundational motor generators

brainstem

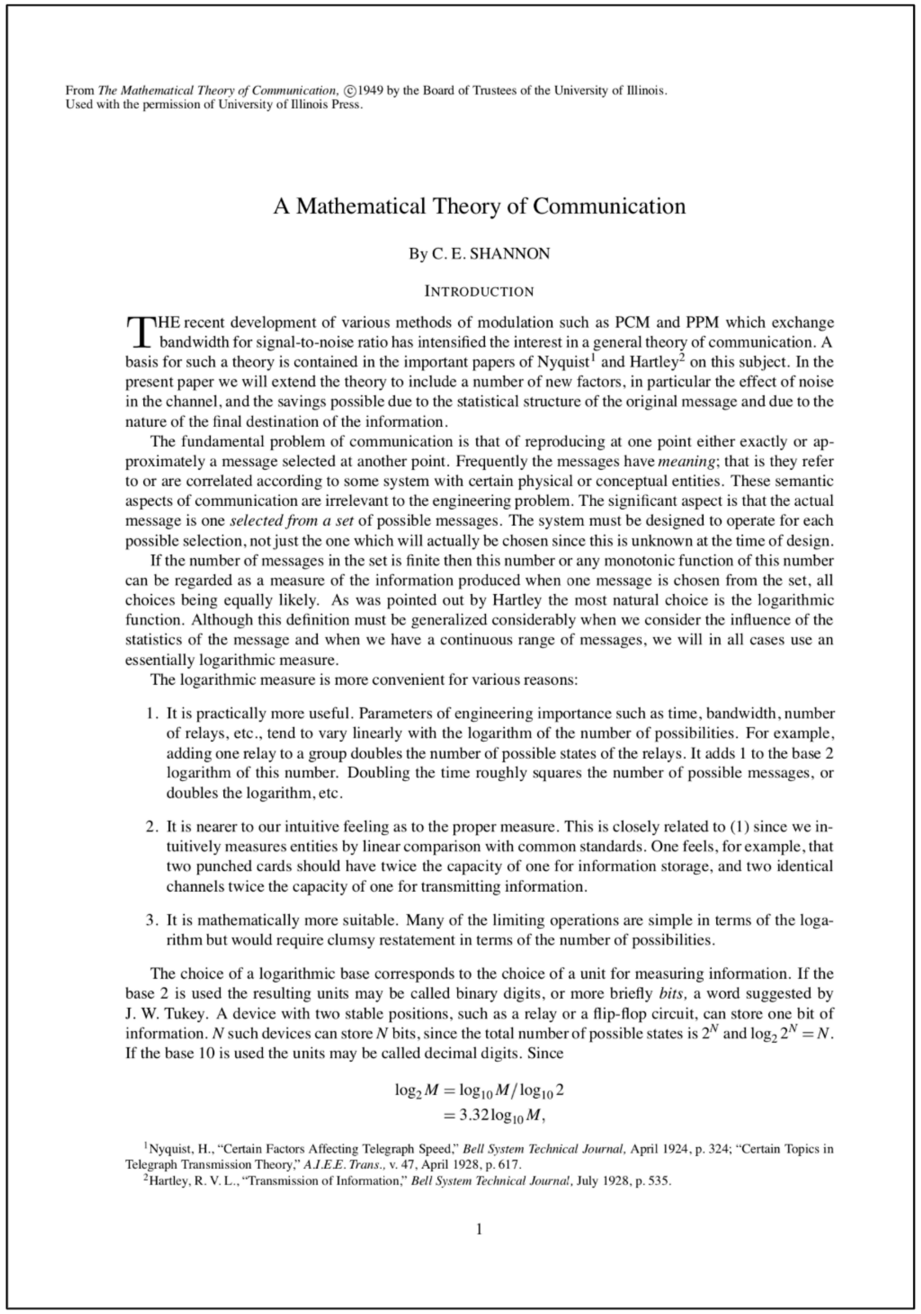
spine

somatic motor output

somatic sensory input

Josh Merel, Matthew Botvinick, and Gregory Wayne. Hierarchical motor control in mammals and machines. Nature Communications, 10(1):5489, December, 2019.

Shannon, Turing, Gödel & von Neumann: The Digital Abstraction†



A Mathematical Theory of Communication

By C. E. SHANNON

Introduction

THE recent development of various methods of modulation such as PCM and PPM which exchange bandwidth for signal-to-noise ratio has intensified the interest in a general theory of communication. A basis for such a theory is contained in the important papers of Nyquist[1] and Hartley[2] on this subject. In the present paper we will extend the theory to include a number of new factors, in particular the effect of noise in the channel, and the savings possible due to the statistical structure of the original message and due to the nature of the final destination of the information.

The fundamental problem of communication is that of reproducing at one point either exactly or approximately a message selected at another point. Frequently the messages have *meaning*; that is they refer to or are correlated according to some system with certain physical or conceptual entities. These semantic aspects of communication are irrelevant to the engineering problem. The significant aspect is that the actual message is one *selected from a set* of possible messages. The system must be designed to operate for each possible selection, not just the one which will actually be chosen since this is unknown at the time of design.

If the number of messages in the set is finite then this number or any monotonic function of this number can be regarded as a measure of the information produced when one message is chosen from the set, all choices being equally likely. As was pointed out by Hartley the most natural choice is the logarithmic function. Although this definition must be generalized considerably when we consider the influence of the statistics of the message and when we have a continuous range of messages, we will in all cases use an essentially logarithmic measure.

The logarithmic measure is more convenient for various reasons:

1. It is practically more useful. Parameters of engineering importance such as time, bandwidth, number of relays, etc., tend to vary linearly with the logarithm of the number of possibilities. For example, adding one relay to a group doubles the number of possible states of the relays. It adds 1 to the base 2 logarithm of this number. Doubling the time roughly squares the number of possible messages, or doubles the logarithm, etc.

2. It is nearer to our intuitive feeling as to the proper measure. This is closely related to (1) since we intuitively measures entities by linear comparison with common standards. One feels, for example, that two punched cards should have twice the capacity of one for information storage, and two identical channels twice the capacity of one for transmitting information.

3. It is mathematically more suitable. Many of the limiting operations are simple in terms of the logarithm but would require clumsy restatement in terms of the number of possibilities.

The choice of a logarithmic base corresponds to the choice of a unit for measuring information. If the base 2 is used the resulting units may be called binary digits, or more briefly *bits,* a word suggested by J. W. Tukey. A device with two stable positions, such as a relay or a flip-flop circuit, can store one bit of information. N such devices can store N bits, since the total number of possible states is 2^N and $\log_2 2^N = N$. If the base 10 is used the units may be called decimal digits. Since

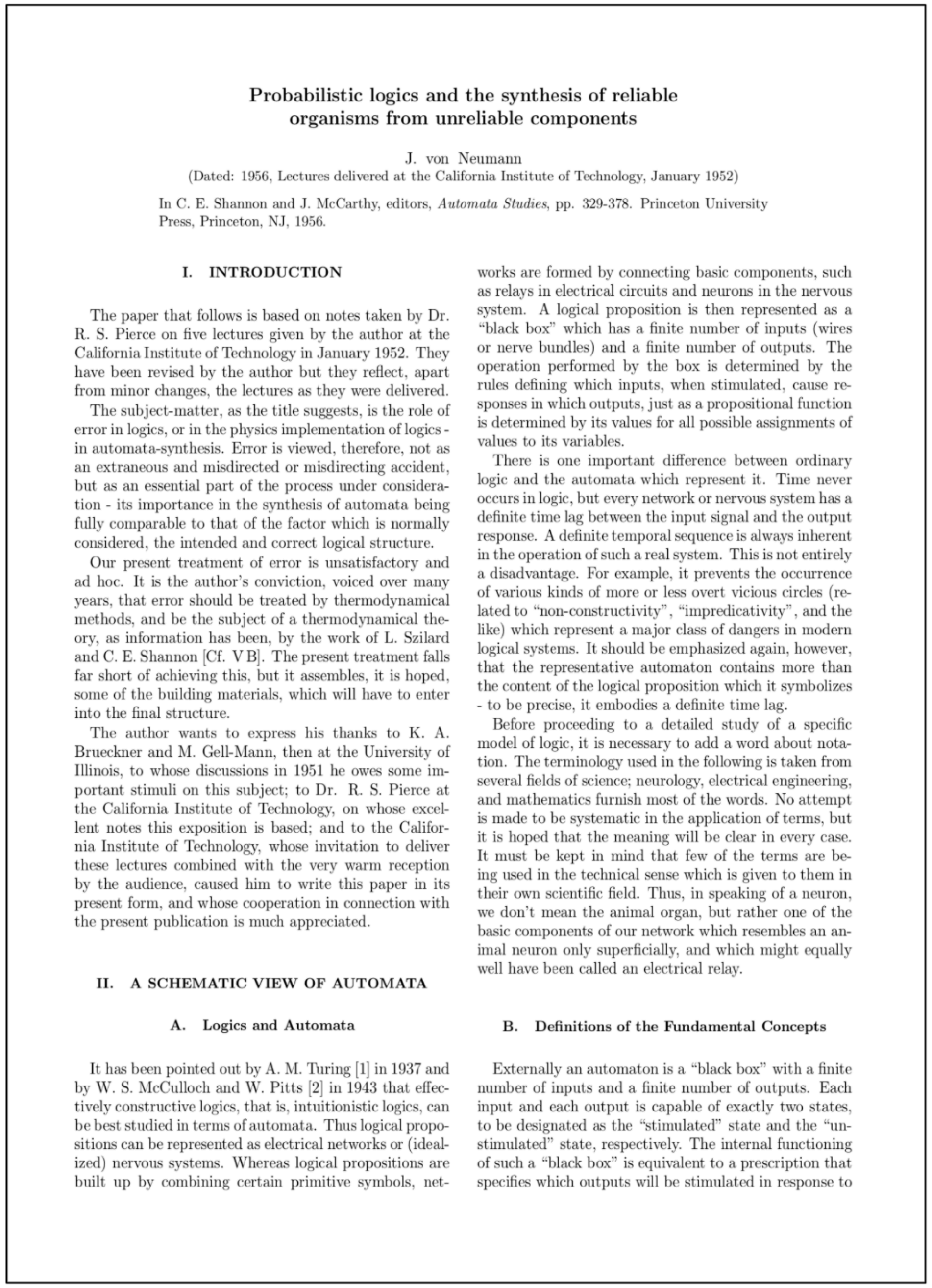
$$\begin{aligned}\log_2 M &= \log_{10} M/\log_{10} 2\\ &= 3.32\log_{10} M,\end{aligned}$$

[1] Nyquist, H., "Certain Factors Affecting Telegraph Speed," *Bell System Technical Journal,* April 1924, p. 324; "Certain Topics in Telegraph Transmission Theory," *A.I.E.E. Trans.,* v. 47, April 1928, p. 617.

[2] Hartley, R. V. L., "Transmission of Information," *Bell System Technical Journal,* July 1928, p. 535.

Claude Shannon

Alan Turing
Kurt Gödel



Probabilistic logics and the synthesis of reliable organisms from unreliable components

J. von Neumann
(Dated: 1956, Lectures delivered at the California Institute of Technology, January 1952)

In C. E. Shannon and J. McCarthy, editors, *Automata Studies*, pp. 329-378. Princeton University Press, Princeton, NJ, 1956.

I. INTRODUCTION

The paper that follows is based on notes taken by Dr. R. S. Pierce on five lectures given by the author at the California Institute of Technology in January 1952. They have been revised by the author but they reflect, apart from minor changes, the lectures as they were delivered.

The subject-matter, as the title suggests, is the role of error in logics, or in the physics implementation of logics - in automata-synthesis. Error is viewed, therefore, not as an extraneous and misdirected or misdirecting accident, but as an essential part of the process under consideration - its importance in the synthesis of automata being fully comparable to that of the factor which is normally considered, the intended and correct logical structure.

Our present treatment of error is unsatisfactory and ad hoc. It is the author's conviction, voiced over many years, that error should be treated by thermodynamical methods, and be the subject of a thermodynamical theory, as information has been, by the work of L. Szilard and C. E. Shannon [Cf. V B]. The present treatment falls far short of achieving this, but it assembles, it is hoped, some of the building materials, which will have to enter into the final structure.

The author wants to express his thanks to K. A. Brueckner and M. Gell-Mann, then at the University of Illinois, to whose discussions in 1951 he owes some important stimuli on this subject; to Dr. R. S. Pierce at the California Institute of Technology, on whose excellent notes this exposition is based; and to the California Institute of Technology, whose invitation to deliver these lectures combined with the very warm reception by the audience, caused him to write this paper in its present form, and whose cooperation in connection with the present publication is much appreciated.

II. A SCHEMATIC VIEW OF AUTOMATA

A. Logics and Automata

It has been pointed out by A. M. Turing [1] in 1937 and by W. S. McCulloch and W. Pitts [2] in 1943 that effectively constructive logics, that is, intuitionistic logics, can be best studied in terms of automata. Thus logical propositions can be represented as electrical networks or (idealized) nervous systems. Whereas logical propositions are built up by combining certain primitive symbols, networks are formed by connecting basic components, such as relays in electrical circuits and neurons in the nervous system. A logical proposition is then represented as a "black box" which has a finite number of inputs (wires or nerve bundles) and a finite number of outputs. The operation performed by the box is determined by the rules defining which inputs, when stimulated, cause responses in which outputs, just as a propositional function is determined by its values for all possible assignments of values to its variables.

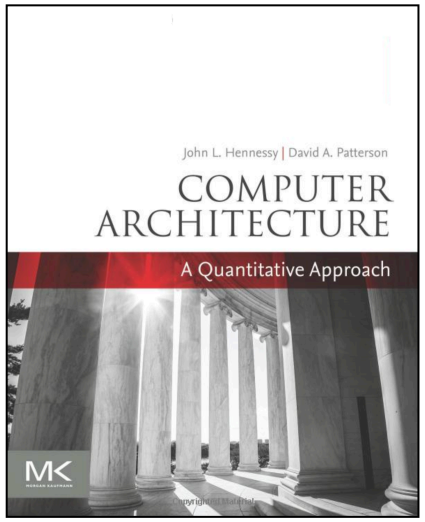
There is one important difference between ordinary logic and the automata which represent it. Time never occurs in logic, but every network or nervous system has a definite time lag between the input signal and the output response. A definite temporal sequence is always inherent in the operation of such a real system. This is not entirely a disadvantage. For example, it prevents the occurrence of various kinds of more or less overt vicious circles (related to "non-constructivity", "impredicativity", and the like) which represent a major class of dangers in modern logical systems. It should be emphasized again, however, that the representative automaton contains more than the content of the logical proposition which it symbolizes - to be precise, it embodies a definite time lag.

Before proceeding to a detailed study of a specific model of logic, it is necessary to add a word about notation. The terminology used in the following is taken from several fields of science; neurology, electrical engineering, and mathematics furnish most of the words. No attempt is made to be systematic in the application of terms, but it is hoped that the meaning will be clear in every case. It must be kept in mind that few of the terms are being used in the technical sense which is given to them in their own scientific field. Thus, in speaking of a neuron, we don't mean the animal organ, but rather one of the basic components of our network which resembles an animal neuron only superficially, and which might equally well have been called an electrical relay.

B. Definitions of the Fundamental Concepts

Externally an automaton is a "black box" with a finite number of inputs and a finite number of outputs. Each input and each output is capable of exactly two states, to be designated as the "stimulated" state and the "unstimulated" state, respectively. The internal functioning of such a "black box" is equivalent to a prescription that specifies which outputs will be stimulated in response to

John von Neumann



Hennessy & Patterson

Connections:
It’s the Network Dummy†

Is there an analog of the Digital Abstraction for modeling the human brain?

Artificial Intelligence and the History of Connectionism

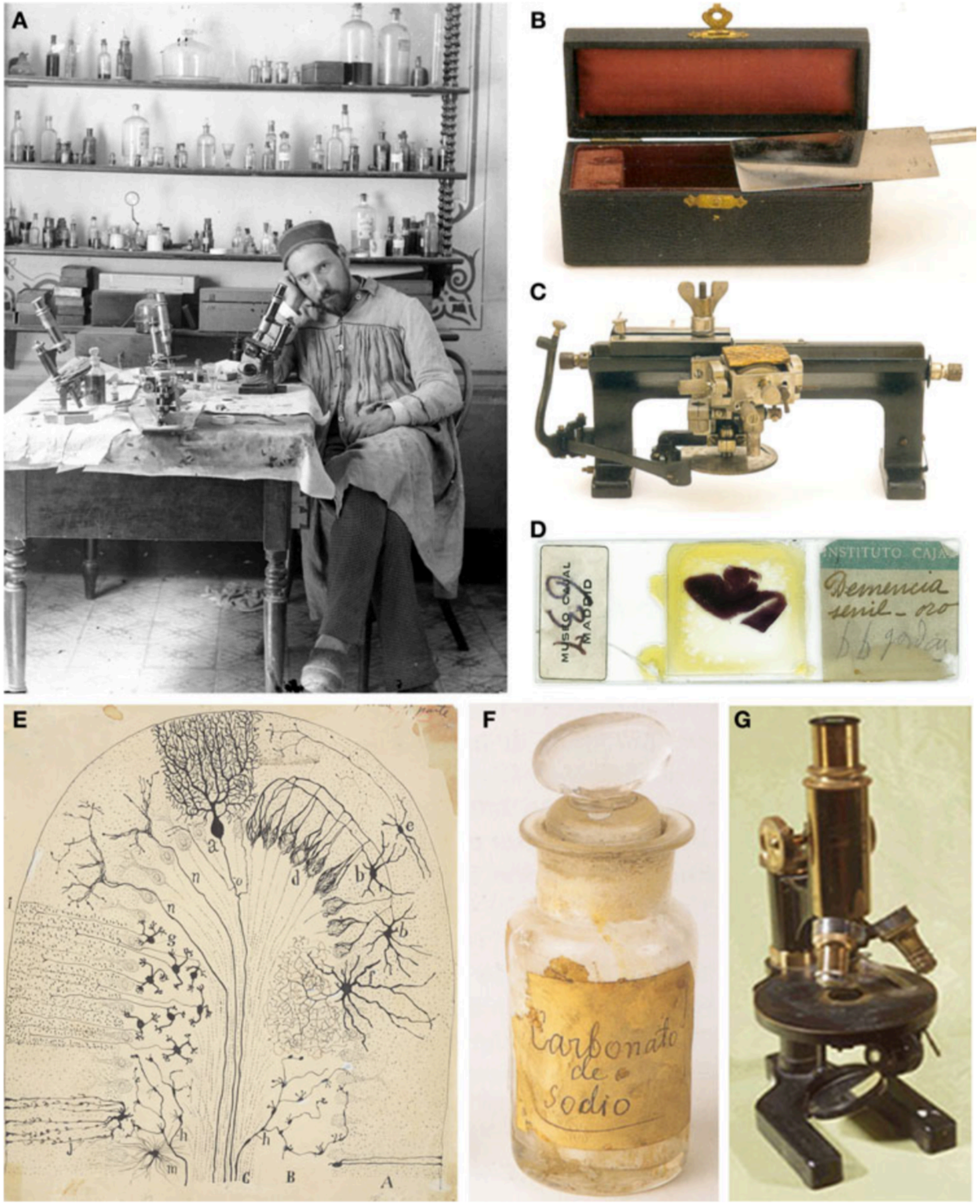
SYMBOLIC	CONNECTIONIST
Fodor & Pylyshyn (LoT)	Rumelhart, McClelland & Hinton (PDP)
COMBINATORIAL	DISTRIBUTED
COMPOSITIONAL	DIFFERENTIABLE
SERIAL	PARALLEL

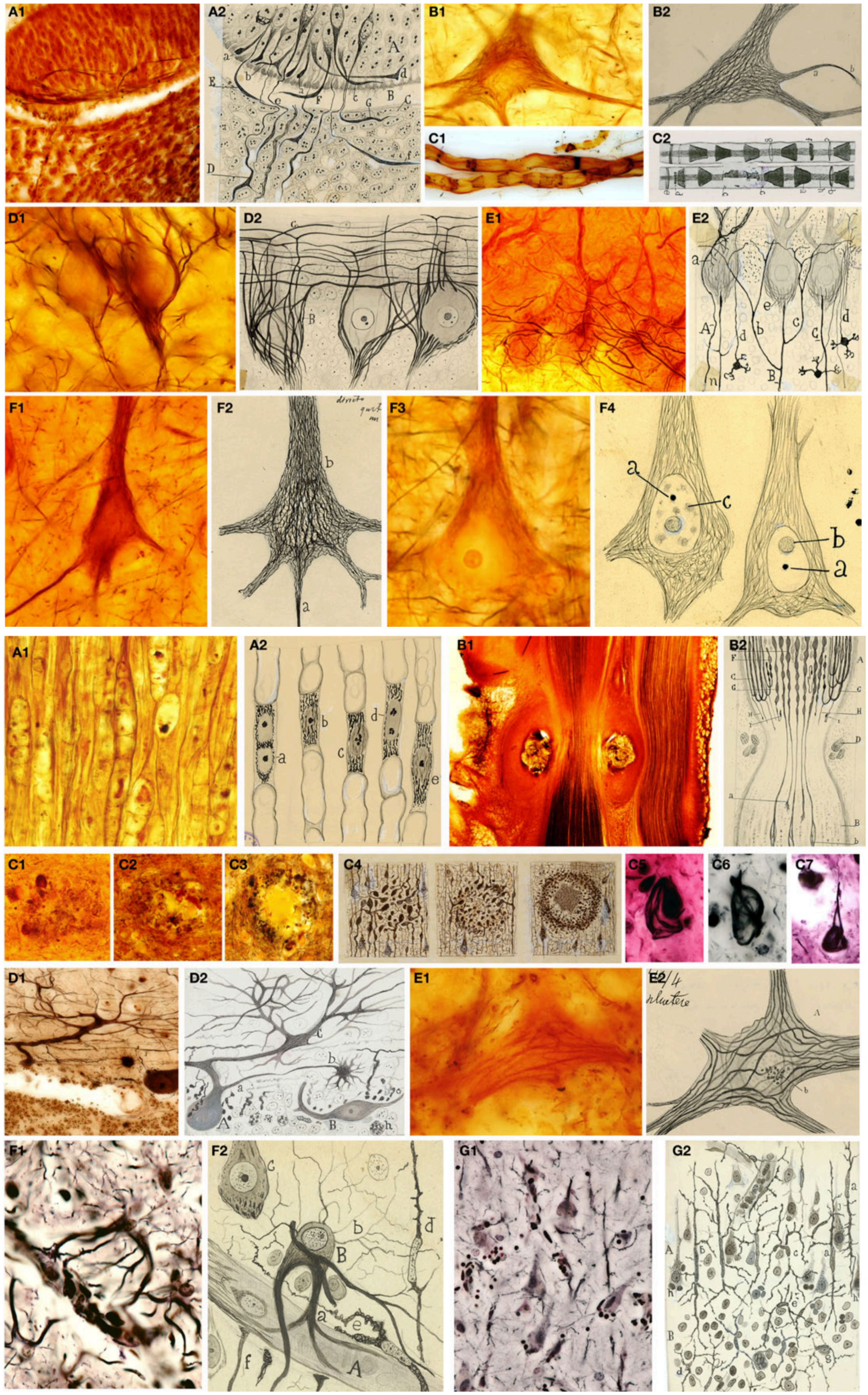
Jerry A. Fodor. *The Language of Thought*. Harvard University Press, Cambridge, MA, 1975.

Jerry A. Fodor and Zenon W. Pylyshyn. Connectionism and cognitive architecture. *Cognition*, 28(1-2):3-71, 1988.

G. E. Hinton, J. L. McClelland, and D. E. Rumelhart. Chapter 3: Distributed Representations. In D. E. Rumelhart and J. L. McClelland, editors, *Parallel Distributed Processing, Explorations in the Microstructure of Cognition: Foundations*. MIT Press, Cambridge, MA, 1986.

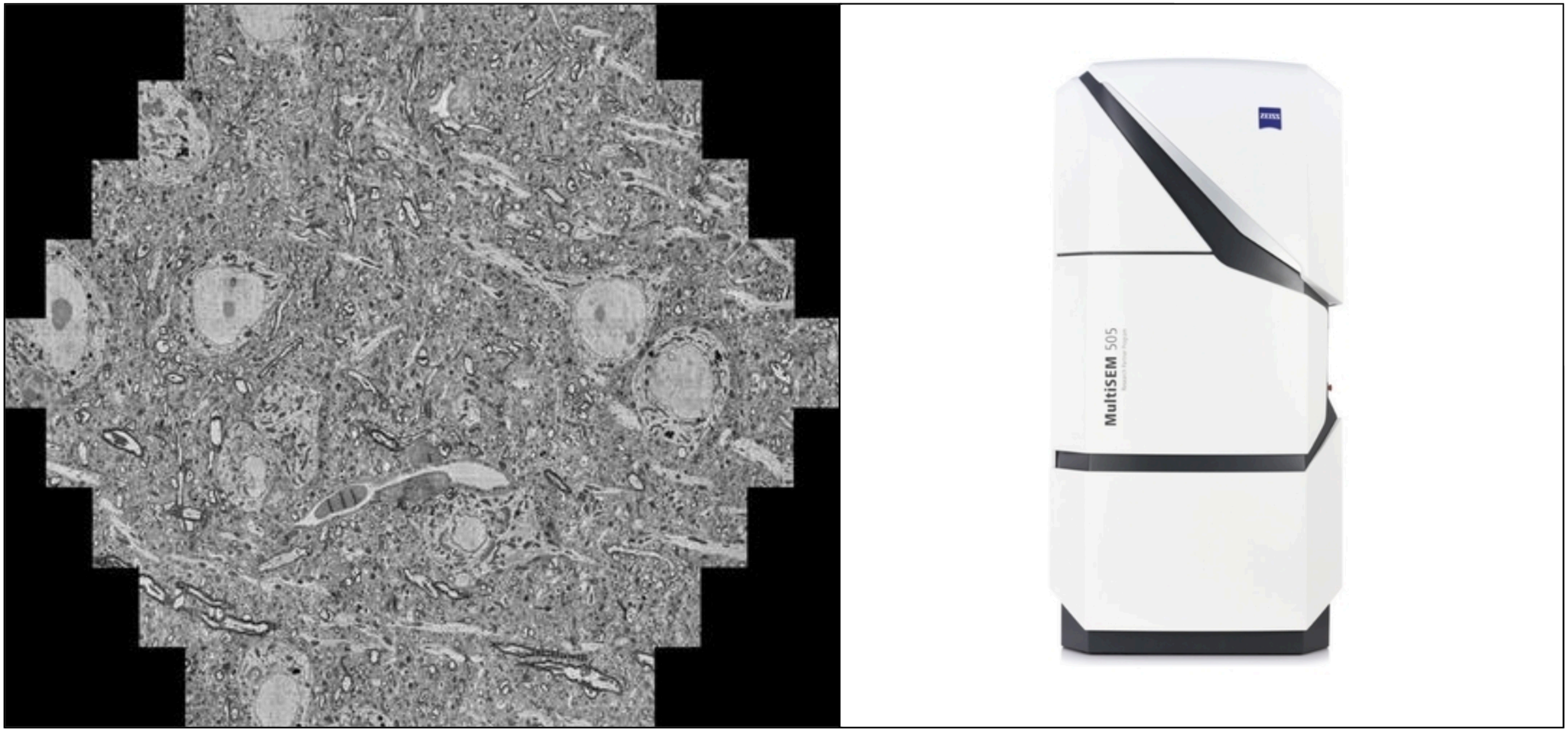
Randall C. O'Reilly, Alex A. Petrov, Jonathan D. Cohen, Christian J. Lebiere, Seth A. Herd, and Trent Kriete. How limited systematicity emerges: A computational cognitive neuroscience approach. In Paco Calvo and John Symons, editors, *The Architecture of Cognition*, pages 191-224. MIT Press, Cambridge, Massachusetts, 2014.





Ortwin Bock. Santiago Ramón y Cajal, Camillo Golgi, Edward Schäfer and the Neuron Doctrine. *Endeavour*. 37(4):228–234, 2013.
Garcia-Lopez, Garcia-Marin, Miguel Freire. The histological slides and drawings of Cajal. *Frontiers in Neuroanatomy*, 4:1-16, 2010.

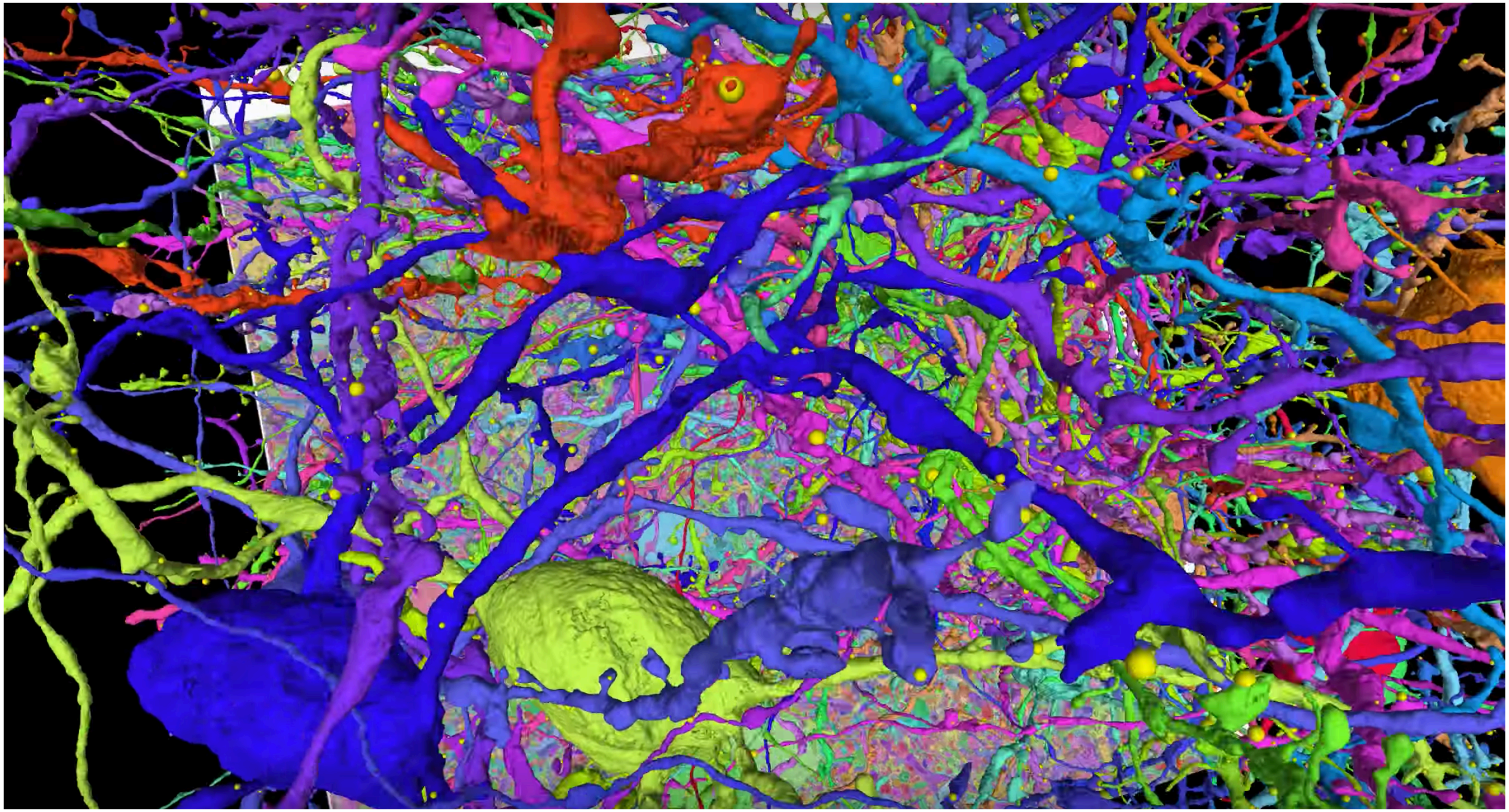
Zeiss Multi-Beam Scanning Electron Microscope



High throughput multi-beam EM imaging with up to 91 parallel beams operating at 2 terapixels per hour with 3.5 nm resolution or better

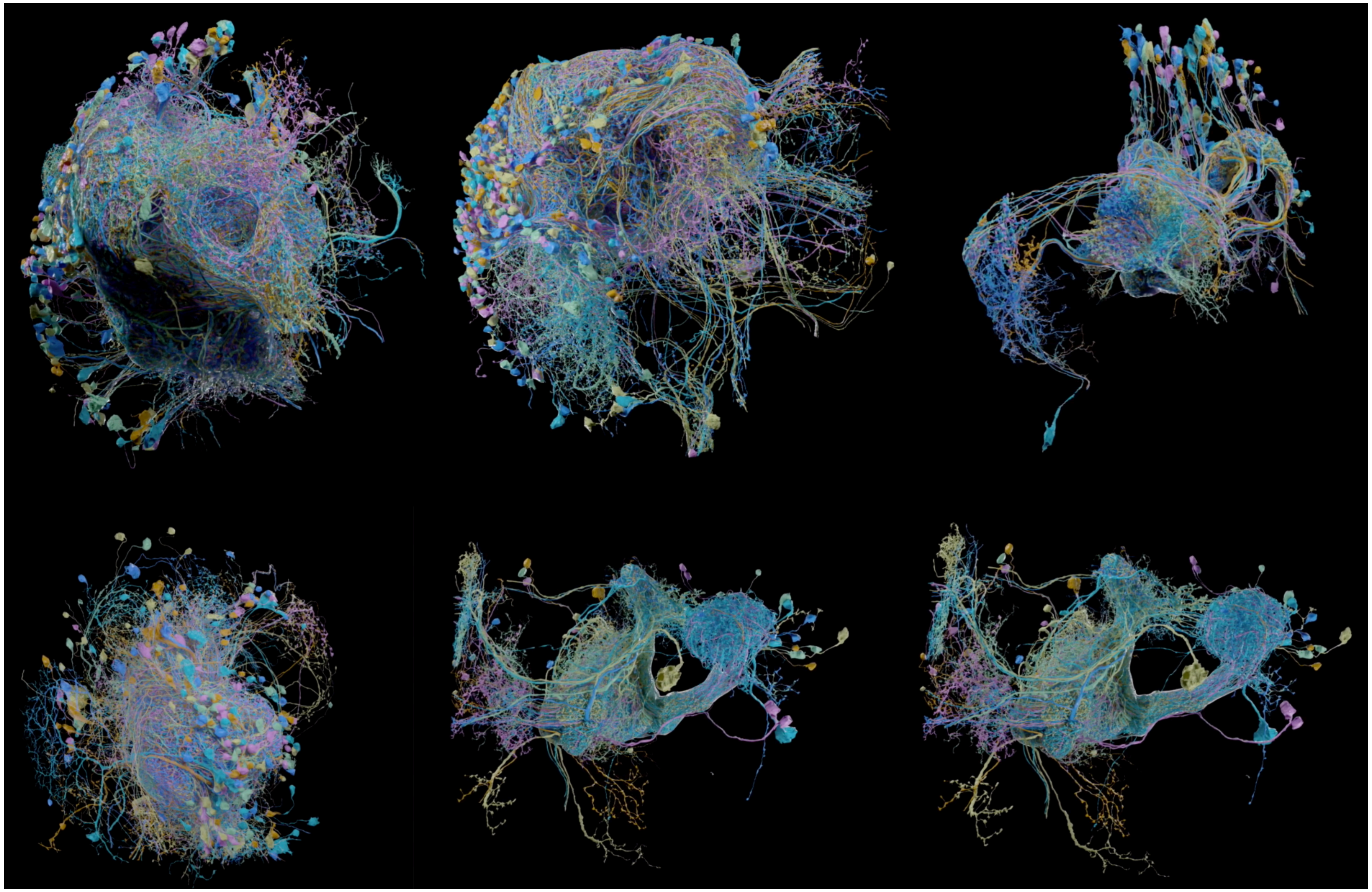
Thomas Dean, Biafra Ahanonu, Mainak Chowdhury, Anjali Datta, Andre Esteva, Daniel Eth, Nobie Redmon, Oleg Rumyantsev, and Ysis Tarter. On the technology prospects and investment opportunities for scalable neuroscience. *CoRR*, arXiv:1307.7302, 2013.

Zebra Finch



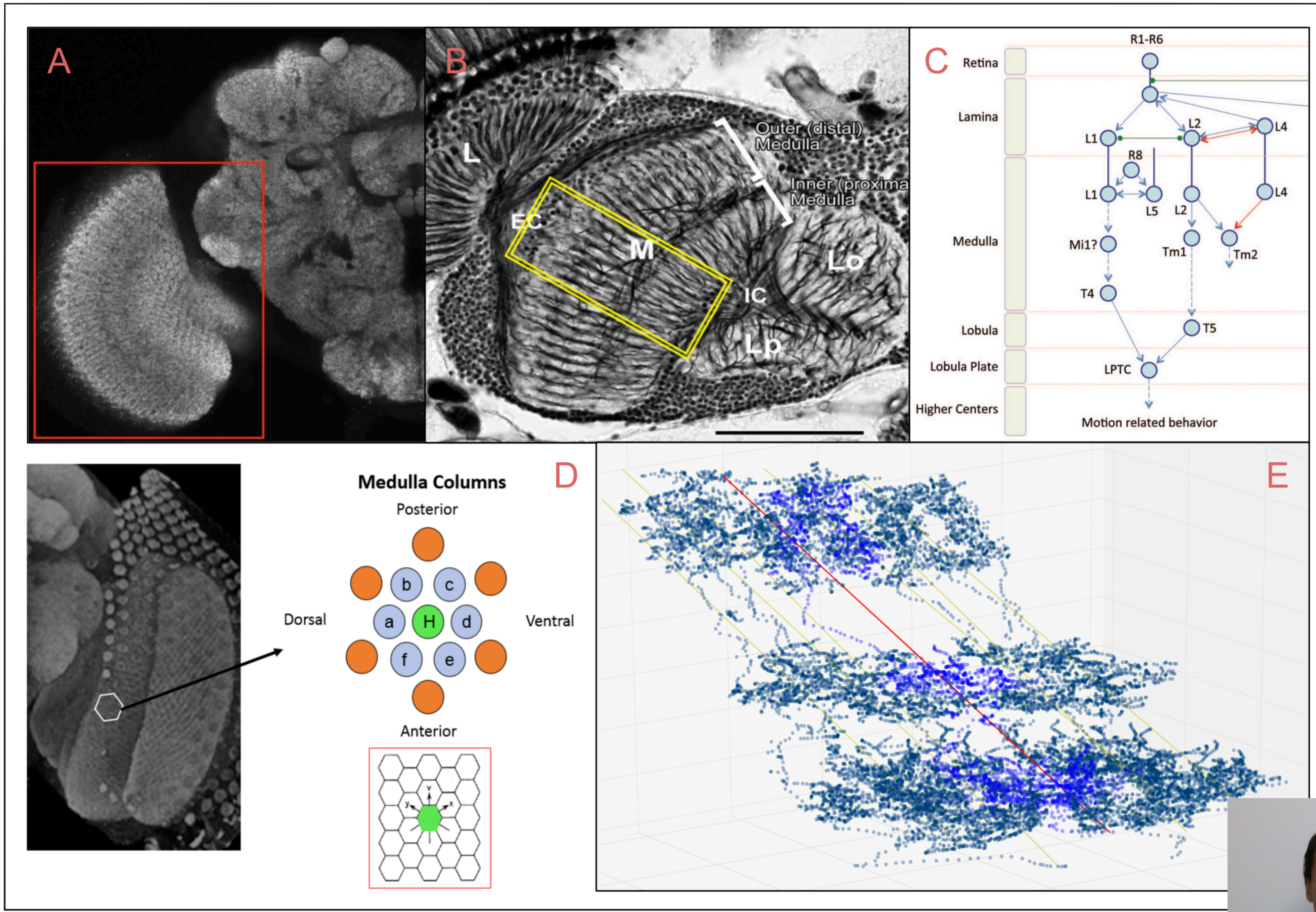
Michal Januszewski, Jörgen Kornfeld, Peter H Li, Art Pope, Tim Blakely, Larry Lindsey, Jeremy B Maitin-Shepard, Mike Tyka, Winfried Denk, and Viren Jain. High-precision automated reconstruction of neurons with flood-filling networks. *Nature Methods*, 15:605--610, 2017 VIDEO

Drosophila Melangaster



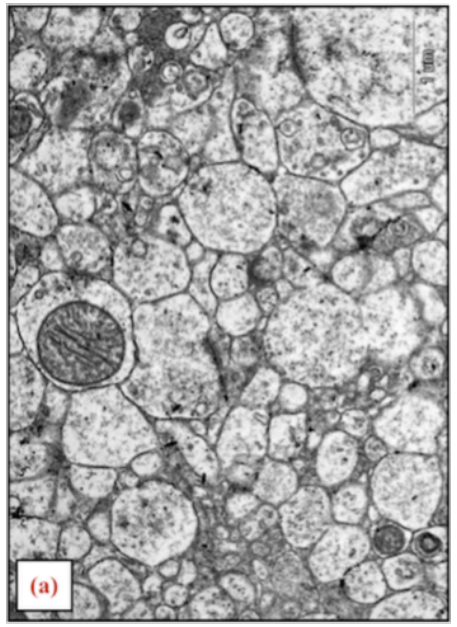
C. Shan Xu, Michal Januszewski, Zhiyuan Lu, Shin-ya Takemura, Kenneth J. Hayworth, Patricia K. Rivlin, Vivek Jayaraman, […], Gerald M. Rubin, Harald F. Hess, Louis K. Scheffer, Viren Jain, and Stephen M. Plaza. A connectome of the adult drosophila central brain. *bioRxiv*, 2020. VIDEO

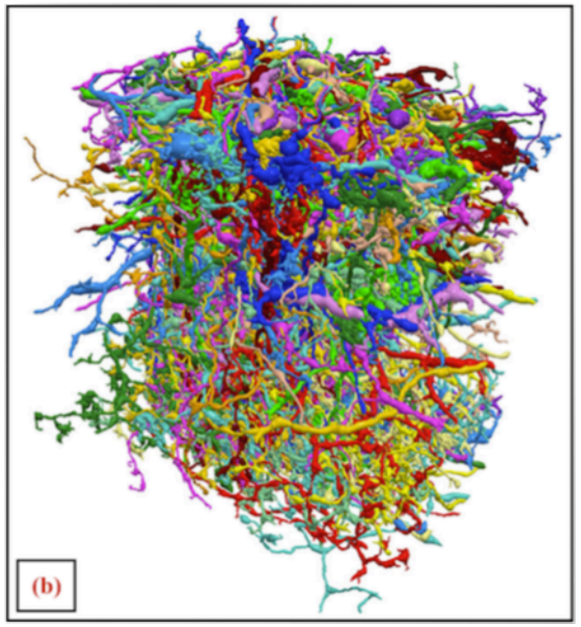
Drosophila Melangaster

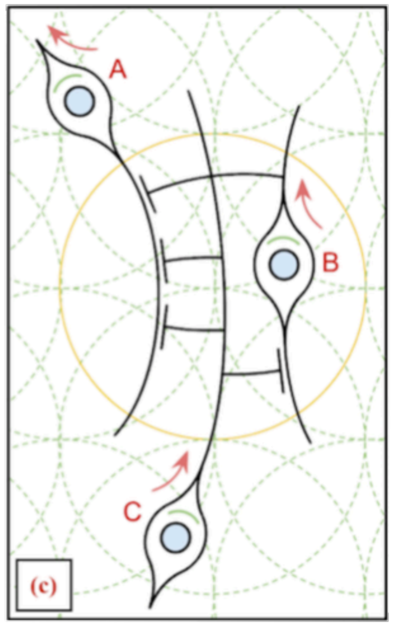


Sophie Aimon, Takeo Katsuki, Logan Grosenick, Michael Broxton, Karl Deisseroth, and Ralph J. Greenspan. Activity sources from fast large-scale brain recordings in adult drosophila. *bioRxiv*, 2015.

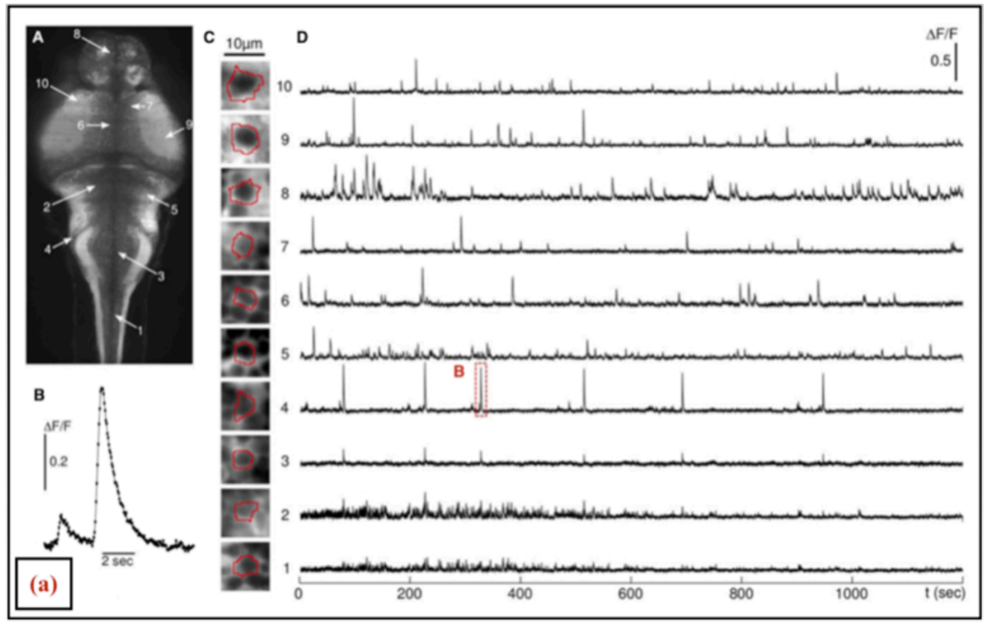
Mesoscale Modeling

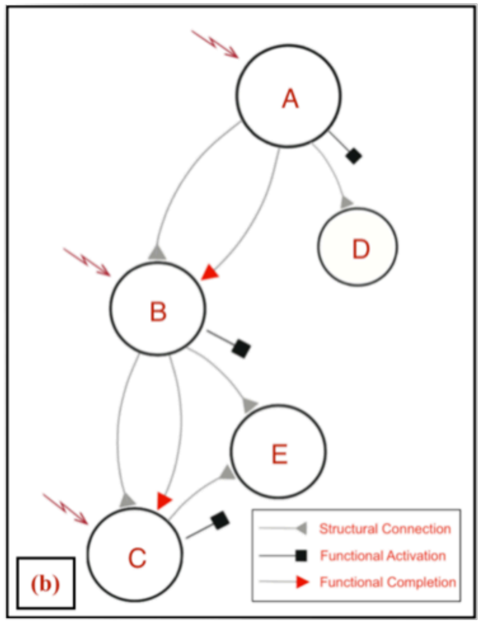






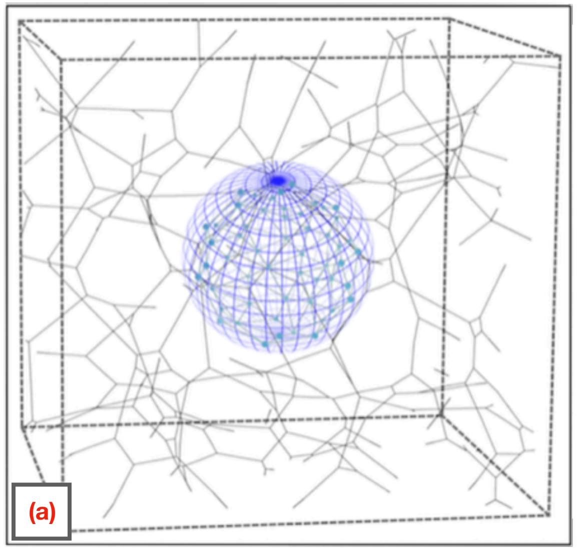
1.(a) serial section sample
1.(b) dense reconstruction
1.(c) synaptic elaborations

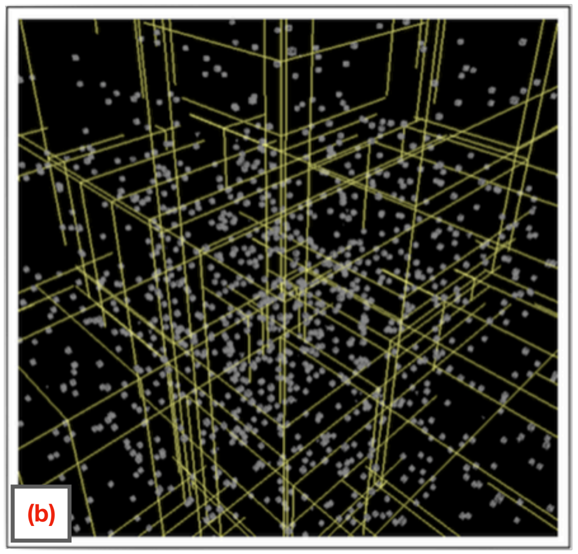


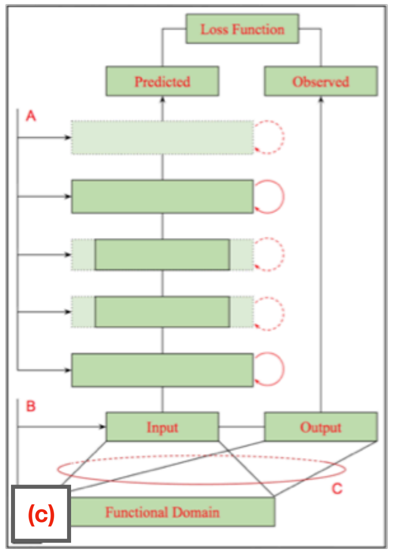


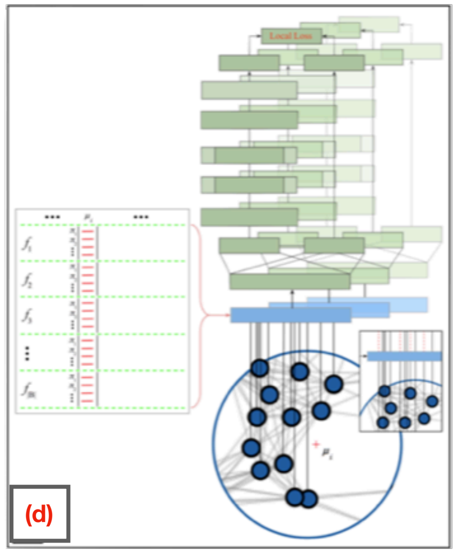
2.(a) dense 2PE recording
2.(b) synaptic transmission

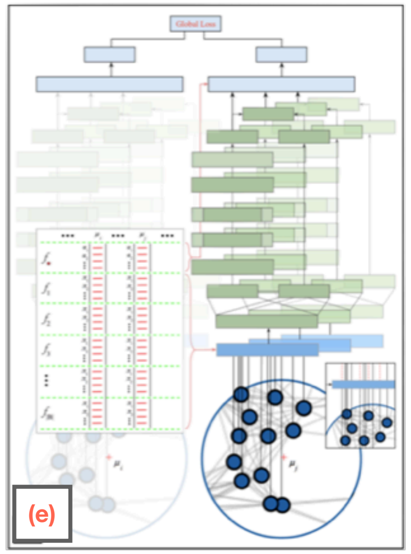
3.(a) 3D embedding space
3.(b) KD tree NN database
3.(c) configurable network
3.(d) basis function filters
3.(e) local and global loss











Thomas Dean. Inferring mesoscale models of neural computation. *CoRR*, arXiv:1710.05183, 2017.