

1

Introduction

The effortless ability of animal brains to engage with their world provides a constant challenge for technology. Despite vast progress in digital computer hardware, software, and system concepts, it remains true that brains far outperform technological computers across a wide spectrum of tasks, particularly when these are considered in the light of power consumption. For example, the honeybee demonstrates remarkable task, navigational, and social intelligence while foraging for nectar, and achieves this performance using less than a million neurons, burning less than a milliwatt, using ionic device physics with a bulk mobility that is about 10 million times lower than that of electronics. This performance is many orders of magnitude more task-competent and power-efficient than current neuronal simulations or autonomous robots. For example, a 2009 ‘cat-scale’ neural simulation on a supercomputer simulated 10^{13} synaptic connections at 700 times slower than real time, while burning about 2 MW (Ananthanarayanan et al. 2009); and the DARPA Grand Challenge robotic cars drove along a densely GPS-defined path, carrying over a kilowatt of sensing and computing power (Thrun et al. 2007).

Although we do not yet grasp completely nature’s principles for generating intelligent behavior at such low cost, neuroscience has made substantial progress toward describing the components, connection architectures, and computational processes of brains. All of these are remarkably different from current technology. Processing is distributed across billions of elementary units, the neurons. Each neuron is wired to thousands of others, receiving input through specialized modifiable connections, the synapses. The neuron collects and transforms this input via its tree-like dendrites, and distributes its output via tree-like axons. Memory instantiated through the synaptic connections between neurons is co-localized with processing through their spatial arrangements and analog interactions on the neurons’ input dendritic trees. Synaptic plasticity is wonderfully complex, yet allows animals to retain important memories over a lifetime while learning on the time scale of milliseconds. The output axons convey asynchronous spike events to their many targets via complex arborizations. In the neocortex the majority of the targets are close to the source neuron, indicating that network processing is strongly localized, with relatively smaller bandwidth devoted to long-range integration.

The various perceptual, cognitive, and behavioral functions of the brain are systematically organized across the space of the brain. Nevertheless at least some aspects of these various

Event-Based Neuromorphic Systems, First Edition.

Edited by Shih-Chii Liu, Tobi Delbruck, Giacomo Indiveri, Adrian Whatley, and Rodney Douglas.

© 2015 John Wiley & Sons, Ltd. Published 2015 by John Wiley & Sons, Ltd.

processes can be discerned within each specialized area, and their organization suggests a coalition of richly intercommunicating specialists. Overall then, the brain is characterized by vast numbers of processors, with asynchronous message passing on a vast point-to-point wired communication infrastructure. Constraints on the construction and maintenance of this wiring enforce a strategy of local collective specialization, with longer range coordination.

For the past two decades neuromorphic engineers have grappled with the implementation of these principles in integrated circuits and systems. The opportunity of this challenge is the realization of a technology for computing that combines the organizing principles of the nervous system with the superior charge carrier mobility of electronics. This book provides some insights and many practical details into the ongoing work toward this goal. These results become ever more important for more mainstream computing, as limits on component density force ever more distributed processing models.

The origin of this neuromorphic approach dates from the 1980s, when Carver Mead's group at Caltech came to understand that they would have to emulate the brain's style of communication if they were to emulate its style of computation. These early developments continued in a handful of laboratories around the world, but more recently there has been an increase of development both in academic and industrial labs across North America, Europe, and Asia. The relevance of the neuromorphic approach to the broader challenges of computation is now clearly recognized (Hof 2014). Progress in neuromorphic methods has been facilitated by the strongly cooperative community of neuroscientists and engineers interested in this field. That cooperation has been promoted by practical workshops such as the Telluride Neuromorphic Cognition Engineering Workshop in the United States, and the CapoCaccia Cognitive Neuromorphic Engineering Workshop in Europe.

Event-Based Neuromorphic Systems arose from this community's wish to disseminate state-of-the-art techniques for building neuromorphic electronic systems that sense, communicate, compute, and learn using asynchronous event-based communication. This book complements the introductory textbook (Liu et al. 2002) that explained the basic circuit building blocks for neuromorphic engineering systems. *Event-Based Neuromorphic Systems* now shows how those building blocks can be used to construct complete systems, with a primary focus on the hot field of event-based neuromorphic systems. The systems described in this book include sensors and neuronal processing circuits that implement models of the nervous systems. Communication between the modules is based on the crucial asynchronous event-driven protocol called the address-event representation (AER), which transposes the communication of spike events on slow point-to-point axons, into digital communication of small data packets on fast buses (see, for example, Chapter 2). The book as a whole describes the state of the art in the field of neuromorphic engineering, including the building blocks necessary for constructing complete neuromorphic chips and for solving the technological challenges necessary to make multi-chip scalable systems. A glance at the index shows the wide breadth of topics, for example, next to 'Moore's law' is 'motion artifact' and next to 'bistable synapse' is 'bootstrapped mirror.'

The book is organized into two parts: Part I (Chapters 2–6) is accessible to readers from a wider range of backgrounds. It describes the range of AER communication architectures, AER sensors, and electronic neural models that are being constructed without delving exhaustively into the underlying technological details. Several of these chapters also include a historical tree that helps relate the architectures and circuits to each other, and that guides readers to the extensive literature. It also includes the largely theoretical Chapter 6 on learning in event-based systems.

Part II (Chapters 7–16) is addressed to readers who intend to construct neuromorphic electronic systems. These readers are assumed to be familiar with transistor physics (particularly subthreshold operation), and in general to be comfortable with reasoning about analog CMOS circuits. A mixed-signal CMOS designer should be comfortable reading these more specialized topics, while an application engineer would be able easily to follow the chapters on hardware and software infrastructure. This part of the book provides information about the various approaches used to construct the building blocks for the sensors and computational units modeling the nervous system, including details of silicon neurons, silicon synapses, silicon cochlea circuits, floating-gate circuits, and programmable digital bias generators. It also includes chapters on hardware and software communication infrastructure and algorithmic processing of event-based sensor output.

The book concludes with Chapter 17, which considers differences between current computers and nervous systems in the ways that computational processing is implemented, and discusses the long-term route toward more cognitive neuromorphic systems.

1.1 Origins and Historical Context

Many of the authors of *Event-Based Neuromorphic Systems* were strongly influenced by *Analog VLSI and Neural Systems* (Mead 1989). Carver Mead's book was the story of an extended effort to apply the subthreshold transistor operating region of CMOS electronics to realize a neural style and scale of computation. The book was written at a time when automatically compiled synchronous logic circuits were just beginning to dominate silicon production, a field that Mead was central in creating. Much like the famous Mead and Conway (1980) book on logic design, which was focused toward instilling a set of methodologies for practical realization of logic chips in digital designers, *Analog VLSI and Neural Systems* was focused on providing a set of organizing principles for neuromorphic designers. These ideas were centered around the name of Mead's group at Caltech, the Physics of Computation group, and emphasized notions such as signal aggregation by current summing on wires, multiplication by summed exponentials, and relations between the fundamental Boltzmann physics of energy barriers and the physics of activation of voltage-sensitive nerve channels.

However, at that time the field was so new that there were many practical aspects that did not work out in the long run, mainly because they suffered from transistor mismatch effects. So the early systems were good for demonstration but not for real-world application and mass production. The fact that current copying in CMOS is the least precise operation possible to implement in practice, was barely mentioned in the book. This omission led to designs that worked ideally in simulation but functioned poorly in practice. In relation to *Event-Based Neuromorphic Systems*, the central importance of communication of information was not realized until after the book was completed, and so none of the systems described in the book had an AER output; rather the analog information was scanned out serially from the systems described there. Even a later collection of chapters (Mead and Ismail 1989) about Mead-lab systems and Mead's review paper in *Proceedings of the IEEE* (1990) barely touched on communication aspects.

Since 1989 there has been a continued drive to improve the technology of neuromorphic engineering. But to place the progress of neuromorphic engineering in context, we can consider logic, that is, digital chip design. Around 1990, a high-end personal computer had about 8 MB

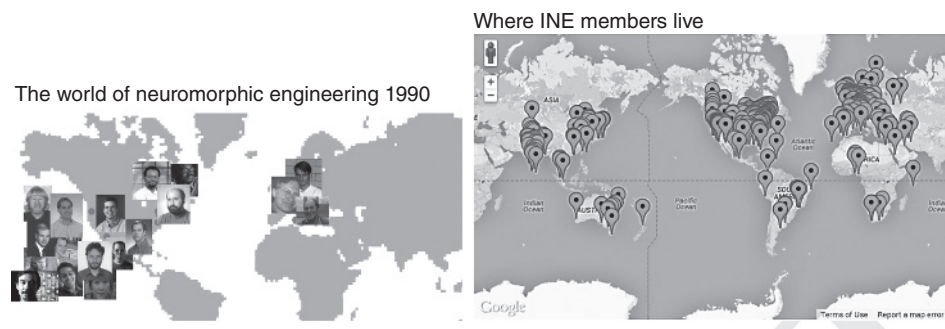


Figure 1.1 Maps of the neuromorphic electrical engineering community in 1990 (left) and 2013 (right, © 2013 Google)

of RAM and about 25 MHz clock speed (one of the authors remembers being a proud owner of a personal CAD station that could be used to work on chip design at home). As of 2013, a state-of-the-art personal computer has about 16 GB of memory and 3 GHz clock speed. So in about 20 years we have seen approximately a 1000-fold increase in memory capacity and a 100-fold increase in clock speed. These of course are reflections of Moore's law and investments of hundreds of billions of dollars. But the basic organizing principles used in computation have hardly changed at all. Most advances have come about because of the availability of more raw memory and computing power, not by fundamental advances in architectures.

During this period the neuromorphic engineering community has expanded considerably from its origins at Caltech, Johns Hopkins, and EPFL (Figure 1.1). At first only a few modest, rather unconvincing lab prototypes could be shown in a couple of labs, and these barely made it off the lab bench. But, after 20 years, neuromorphic engineering has scaled the number of spiking neurons in a system from a few hundred up to about a million (Chapter 16), neuromorphic sensors are available as high-performance computer peripherals (Liu and Delbruck 2010), and these components can be used by people at neuromorphic engineering workshops who know little about transistor-level circuit design (Cap n.d.; Tel n.d.). The literature shows a steady exponential growth in papers with the keywords 'neuromorphic' or 'address-event representation' (Figure 1.2), which is a higher growth rate than for the term 'synchronous logic.' Although the slopes of these exponentials tend to flatten over time, the number of

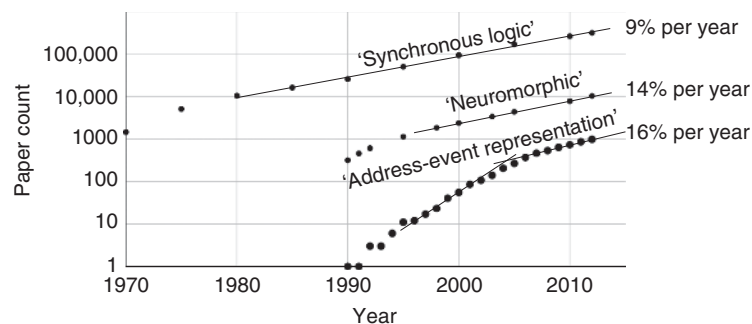


Figure 1.2 Growth of literature over time. From Google Scholar

papers mentioning ‘address-event representation’ has increased for the last 5 years at the rate of about 16% per year. If this growth is considered as resulting from perhaps 15 labs working for an average of 15 years at an investment of \$200,000 per year, then this progress has been achieved at a total financial investment of perhaps 50 million dollars, a tiny fraction of the hundreds of billions spent on developing conventional electronics during this period.

1.2 Building Useful Neuromorphic Systems

To be adopted as mainstream technology and have the financial support and competitive environment of an active industrial market there are some obvious requirements. Neuromorphic systems must function robustly and repeatably across chips, across temperature, with noisy power supplies, must have interfaces that allow easy development of applications, and need to be portable for use in the field without specialized equipment. *Event-Based Neuromorphic Systems* teaches the knowledge of the required technologies built up over the past 20 years of effort.

These features of a neuromorphic electronic system are necessary but not sufficient. A neuromorphic system must outperform conventional technology, or at least justify the investment of effort based on the belief that it could outperform conventional approaches when scaled up or when silicon technology can no longer scale down to smaller feature sizes or power supply voltages. And this last point has been a weakness: Proposals have not convincingly shown that the neuromorphic approach is better than simply scaling logic and making it more parallel. Many grants have been funded nonetheless, but the proposals are too vague to be very credible. One could say that funders are simply hopeful; there is no alternative that offers anything but a new device technology (e.g., graphene) to enable yet more clock and feature size scaling.

The scaling problem brings up the importance of communication: To scale up systems requires growing not only smaller in feature size and cost but also larger in system capabilities. To be neuromorphic, these systems must emulate something like the hybrid data-driven computation and communication architecture used in brains with their massive numbers of connections. One can see the evidence for this requirement from the direction of conventional electronics as well, with logic systems becoming more parallel and distributed. This requirement for communication is why the neuromorphic community has focused its efforts on event-based architectures and it is why this book is aimed at teaching the state-of-the-art techniques for building such systems. Chapter 2 will begin by outlining the principles of event-based communication architectures for neuromorphic systems.

References

- Ananthanarayanan R, Esser SK, Simon HD, and Modha DS. 2009. The cat is out of the bag: cortical simulations with 10^9 neurons, 10^{13} synapses. Proceedings of the Conference on High Performance Computing Networking, Storage and Analysis, Portland, OR, November 14–20, 2009. IEEE. pp. 1–12.
- Cap. n.d. Capo Caccia Cognitive Neuromorphic Engineering Workshop, <http://capocaccia.ethz.ch/> (accessed July 16, 2014).
- Hof RD. 2014. Qualcomm’s neuromorphic chips could make robots and phones more astute about the world. *MIT Technology Review*. <http://www.technologyreview.com/featuredstory/526506/neuromorphic-chips/>.
- Liu SC and Delbruck T. 2010. Neuromorphic sensory systems. *Curr. Opin. Neurobiol.* **20**(3), 288–295.
- Liu SC, Kramer J, Indiveri G, Delbrück T, and Douglas R. 2002. *Analog VLSI: Circuits and Principles*. MIT Press.

- Mead CA. 1989. *Analog VLSI and Neural Systems*. Addison-Wesley, Reading, MA.
- Mead CA. 1990. Neuromorphic electronic systems. *Proc. IEEE* **78**(10), 1629–1636.
- Mead CA and Conway L. 1980. *Introduction to VLSI Systems*. Addison-Wesley, Reading, MA.
- Mead CA and Ismail M (eds). 1989. *Analog VLSI Implementation of Neural Systems*. Kluwer Academic Publishers, Norwell, MA.
- Tel. n.d. Telluride Neuromorphic Cognition Engineering Workshop, www.ine-web.org/ (accessed July 16, 2014).
- Thrun S, Montemerlo M, Dahlkamp H, Stavens D, Aron A, Diebel J, Fong P, Gale J, Halpenny M, Hoffmann G, Lau K, Oakley C, Palatucci M, Pratt V, Stang P, Strohband S, Dupont C, Jendrossek LE, Koelen C, Markey C, Rummel C, Niekerk J, Jensen E, Alessandrini P, Bradski G, Davies B, Ettinger S, Kaehler A, Nefian A, and Mahoney P. 2007. Stanley: the robot that won the DARPA grand challenge. In: *The 2005 DARPA Grand Challenge* (eds Buehler M, Iagnemma K, and Singh S). Vol. 36: Springer Tracts in Advanced Robotics. Springer, Berlin Heidelberg. pp. 1–43.