Tutorial: Large-Scale Spiking Neuromorphic Architecture Exploration using SANA-FE

James A. Boyle*, Mark Plagge†, Suma G. Cardwell†, Frances S. Chance† and Andreas Gerstlauer*
*The University of Texas at Austin, Texas, USA
†Sandia National Laboratories, New Mexico, USA

***Abstract*—Neuromorphic computing uses brain-inspired concepts to accelerate and efficiently execute a wide range of applications, such as mimicking biological circuits, solving NP-hard optimization problems and accelerating machine learning at the edge. In particular, neuromorphic architectures to efficiently execute Spiking Neural Networks (SNNs) have gained popularity. SNNs extend artificial neural networks (ANNs) by encoding information in time as either rates or delays between spiking events, shared between neurons via their weighted connections. SNN-based platforms are event-driven, resulting in naturally sparse, noise-tolerant and power-efficient computation.**

In this tutorial, we present the state-of-the-art in scalable digital and analog spiking neuromorphic system architectures, and discuss current research trends within the neuromorphic architecture field at the system level. We further introduce our SANA-FE tool for Simulation of Advanced Neuromorphic Architectures for Fast Exploration, which has been developed as part of a collaboration between the University of Texas at Austin and Sandia National Laboratories. SANA-FE allows for modeling and performance-power prediction of different spiking hardware architectures executing SNN applications to support rapid, early system-level design-space exploration, hardware-aware application development and system architecture co-design. The tutorial includes a hands-on component in which SANA-FE's capabilities are demonstrated and used to perform system design and application mapping case studies.

I. Introduction

Neuromorphic computing uses brain-inspired concepts to accelerate and efficiently execute a wide range of applications, such as mimicking biological circuits [1]–[3], solving NP-hard optimization problems [4], [5] and accelerating machine learning at the edge [6]. In particular, neuromorphic architectures to efficiently execute Spiking Neural Networks (SNNs) have gained popularity. SNNs extend artificial neural networks (ANNs) by encoding information in time as either rates or delays between spiking events, shared between neurons via their weighted connections. SNN-based platforms are event-driven, resulting in naturally sparse, noise-tolerant and power-efficient computation.

A range of hardware platforms have been proposed for efficiently executing SNNs, varying widely in their design approaches. Both digital and mixed-signal architectures have been implemented, using novel design elements such as custom logic to emulate biological neurons and network architectures optimized for spiking communication patterns. Future architectures will further leverage novel emerging devices to achieve improvements in power and performance for specific applications. However, the design-space for these architectures is large, and developing the next generation of neuromorphic systems will require co-design across applications, architectures, circuits and devices.

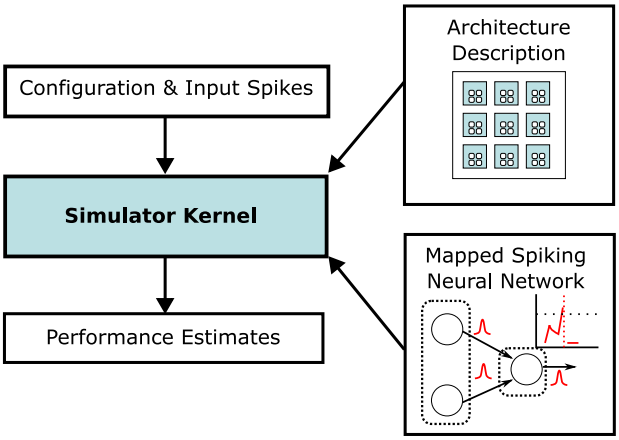

Fig. 1. Overview of SANA-FE.

Architecting new neuromorphic chips involves several design decisions that can affect power performance. Performance models can be used to estimate the impact of different approaches and inform these decisions. SANA-FE (Simulating Advanced Neuromorphic Architectures for Fast Exploration) is an open-source tool developed in a collaboration between The University of Texas at Austin and Sandia National Laboratories to rapidly and accurately model and simulate the energy and performance of different neuromorphic hardware platforms [7], [8]. An overview of SANA-FE is shown in Fig. 1. The simulator takes a description of a hardware platform and SNN mapped onto the hardware to model execution of the SNN and predict power and performance. SANA-FE allows for modeling and performance-power prediction of different spiking hardware architectures executing SNN applications to support rapid, early system-level design-space exploration, hardware-aware application development and system architecture co-design.

In this tutorial, we present the state-of-the-art in scalable digital and analog spiking neuromorphic system architectures, and discuss current research trends within the neuromorphic architecture field at the system level. We further introduce our SANA-FE tool and include a hands-on component in which SANA-FE's capabilities are demonstrated and used to perform system design and application mapping case studies.

The tutorial is organized in two parts as follows: In the first part, we will provide an overview and introduction to large-scale spiking architectures and our neuromorphic hardware simulator SANA-FE. In the second part, we will then switch to the hands-on component including a (1) walk-through of SANA-FE installation and setup using Docker, (2) an overview of SANA-FE's architecture, SNN and output trace file formats, and (3) a demo and mapping-space exploration challenge for Intel's Loihi platform executing a real-world application.

II. Tutorial Overview

In the following, we describe the two tutorial parts covering an introduction to neuromorphic computing and our SANA-FE simulator as well as the hands-on SANA-FE demonstration in more detail.

A. Large-Scale Spiking Neuromorphic Architectures

Our tutorial starts with an introduction to neuromorphic computing, including an overview of existing neuromorphic hardware platforms. We first explore how spiking neural networks (SNNs) can solve different tasks, and how spike events can encode information in time and space. We then describe how custom hardware platforms have been designed to efficiently execute SNNs, and compare different hardware architectures including purely digital platforms such as Intel's Loihi [9] as well as analog realizations such as BrainScaleS-2 [10]. We present emerging trends in the neuromorphic field, including how novel hardware elements may be incorporated into designs to achieve even better efficiency e.g., analog circuits and devices that imitate the dynamics of biological neurons.

Next, we introduce our open-source, architectural-level simulator for Simulating Advanced Neuromorphic Architectures for Fast Exploration (SANA-FE). This introduction describes the various input and output file formats used by SANA-FE, and includes examples of its architecture and SNN description file formats. We explain how the simulator kernel accurately models hardware updates and predicts energy and performance, in a course-grained time-step based loop.

B. SANA-FE Demonstration

The second section is hands-on and include exercises to showcase SANA-FE's capabilities. This includes initial exercises to extend a simple hardware architecture. Then, we modify a small SNN and map it to the extended hardware. Using this architecture and mapped SNN, we show how to generate various spike, neuron and hardware activity traces. We conclude with a demonstration of a larger real-world application, categorizing hand gestures using data from a neuromorphic sensor [11]. Using the scripting capabilities of SANA-FE, we show how SANA-FE's rapid performance estimates can enable effective exploration and aid with co-design.

This demonstration uses a Docker environment (available at [8]) that contains the required binaries and inputs.

Presenter Biographies

James Boyle is a Ph.D. student in the Chandra Department of Electrical and Computer Engineering at The University of Texas at Austin, USA, under the advisement of Professor Andreas Gerstlauer. He received his M.Eng. in electronic engineering with computer systems from the University of Southampton, UK in 2016. His research focuses on computer architecture and neuromorphic computing.

Andreas Gerstlauer is a Cullen Trust for Higher Education Endowed Professor and Associate Chair in the Electrical and Computer Department at The University of Texas at Austin. He received the Ph.D. degree in Information and Computer Science from the University of California at Irvine (UCI), Irvine, CA, USA, in 2004. Prior to joining UT Austin in 2008, he was an Assistant Researcher with UCI. His research interests cover system-level design and embedded systems.

Acknowledgments

This article has been authored by an employee of National Technology & Engineering Solutions of Sandia, LLC under Contract No. DE-NA0003525 with the U.S. Department of Energy (DOE). The employee owns all right, title and interest in and to the article and is solely responsible for its contents. The United States Government retains and the publisher, by accepting the article for publication, acknowledges that the United States Government retains a non-exclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this article or allow others to do so, for United States Government purposes. The DOE will provide public access to these results of federally sponsored research in accordance with the DOE Public Access Plan https://www.energy.gov/downloads/doe-public-access-plan.
SAND-XXXX

References

[1] C. Mead, *Analog VLSI and Neural Systems*. Reading, MA: Addison-Wesley, 1989.
[2] G. Indiveri *et al.*, "Neuromorphic silicon neuron circuits," *Front. Neurosci.*, vol. 5, p. 73, 2011.
[3] C. S. Thakur *et al.*, "Large-scale neuromorphic spiking array processors: A quest to mimic the brain," *Front. Neurosci.*, vol. 12, p. 891, 2018.
[4] J. D. Smith *et al.*, "Neuromorphic scaling advantages for energy-efficient random walk computations," *Nat. Electron.*, vol. 5, no. 2, pp. 102–112, 2022.
[5] J. Aimone *et al.*, "A review of non-cognitive applications for neuromorphic computing," *Neuromorphic Computing and Engineering*, 2022.
[6] C. D. Schuman, S. R. Kulkarni, M. Parsa, J. P. Mitchell, P. Date, and B. Kay, "Opportunities for neuromorphic computing algorithms and applications," *Nat. Comput. Sci.*, vol. 2, no. 1, pp. 10–19, 2022.
[7] J. Boyle, M. Plagge, S. G. Cardwell, F. S. Chance, and A. Gerstlauer, "Performance and energy simulation of spiking neuromorphic architectures for fast exploration," in *Int. Conf. Neuromorphic Syst. (ICONS)*, 2023.
[8] SANA-FE. [Online]. Available: github.com/SLAM-Lab/SANA-FE
[9] M. Davies *et al.*, "Loihi: A neuromorphic manycore processor with on-chip learning," *IEEE Micro*, vol. 38, no. 1, pp. 82–99, 2018.
[10] C. Pehle *et al.*, "The BrainScaleS-2 accelerated neuromorphic system with hybrid plasticity," *Front. Neurosci.*, vol. 16, p. 795876, 2022.
[11] R. Massa, A. Marchisio, M. Martina, and M. Shafique, "An efficient spiking neural network for recognizing gestures with a DVS camera on the Loihi neuromorphic processor," in *Int. Joint Conf. Neural Networks (IJCNN)*, 2020.