

Cortical Columns Computing Systems (C3S): Towards a Silicon Neocortex for AI Compute

Harideep Nair and John Paul Shen

Electrical and Computer Engineering Department, Carnegie Mellon University

Abstract—Reverse-engineering the human neocortex has been a grand challenge in machine learning and computer architecture. Although deep neural networks, born from this challenge, have achieved remarkable success, they resort to intense biologically implausible statistical processing on large datasets resulting in exponentially increasing compute demand. Recent neuroscience research suggests Cortical Columns as the fundamental processing units in the neocortex that encapsulate intelligence. This abstract presents ongoing research focusing on building such *Cortical Columns Computing Systems (C3S)* as a first step towards a silicon neocortex with brain-like capabilities and efficiency. Initial findings indicate that designing extremely energy-efficient intelligent C3S-based sensory processing chiplets, using off-the-shelf CMOS technology, is quite feasible and very promising, and certainly warrants further research exploration.

Index Terms—neuromorphic computing, cortical columns, temporal neural networks, online sensory processing

I. INTRODUCTION AND BACKGROUND

Human brain (neocortex) is highly adept at sensory processing tasks. Current deep neural networks (DNNs) are attempting to replicate such human-like sensory processing but using conventional hardware that employ Turing computation model and von-Neumann computer architecture targeting numerical computation. DNNs have achieved excellent results in visual object recognition, natural language processing, etc. However, the computation required is increasing at an exponential rate. Many believe this trend is not sustainable and an alternative paradigm is needed [9].

Neuromorphic spiking neural networks (SNNs) offer an alternative approach to AI compute, based on brain-inspired principles and brain-like architecture. Temporal Neural Networks (TNNs) [7] emerged recently as a more biologically plausible approach utilizing efficient temporal encoding, local spike timing dependent plasticity learning and neocortical hierarchy comprising of synapses, neurons, (mini)columns, and layers. Further, Hawkins’ new theory on intelligence [2] proposes *Cortical Columns (CCs)* as the fundamental neocortical processing units. Higher forms of intelligence involve result from larger number of CCs. The neocortex gains its intelligence through CC’s ability to model sensory information in structured Reference Frames (RFs), and learn complete models of objects through predict-sense-update feedback loops. There is synergy between CCs and TNNs; CCs resemble TNNs with recurrence, embodied in the RFs.

This paper presents our ongoing research in designing and implementing *Cortical Columns Computing Systems (C3S)*, capable of performing highly energy-efficient online sensory

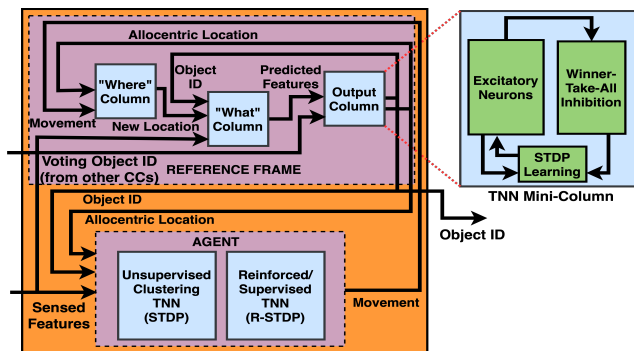


Fig. 1: Each Cortical Column (CC) consists of five TNN-style mini-columns: *Where*, *What* and *Output* mini-columns that together implement the Reference Frame (RF), and *unsupervised* and *supervised* mini-columns comprising the agent. For visual object recognition, the respective functionalities of the three RF mini-columns are: derive locations of sensor on the object, map features to locations, and derive the object ID based on the feature map. Each CC learns continuously through feedback (recurrence) by capturing meaningful patterns in its RF.

processing [6]. Unlike many neuromorphic works that focus on exotic device technologies such as resistive memory, we utilize standard off-the-shelf digital CMOS targeting near-future mass market computing. We envision C3S-based neuromorphic sensory processing units (NSPUs) as specialized chiplets that can be easily integrated into any edge AI compute substrate.

II. CORTICAL COLUMNS COMPUTING SYSTEM

Inspired from Smith’s macrocolumn architecture [8], our proposed Cortical Column (CC) microarchitecture (shown in Figure 1) consists of two components: 1) a *Reference Frame* that maintains a “map” of the sensory information, and 2) an *Agent* that achieves goal-oriented behavior based on information from the Reference Frame and the current input signals. Agent comprises of two TNN-type mini-columns performing unsupervised clustering and supervised classification.

Reference Frame involves three functional modules in the form of three types of mini-columns: *Where*, *What* and *Output*. In the context of visual object recognition, the three types of mini-columns together create models of objects by tracking locations of features on the object. *Output* determines the object identity. *Where* generates location of sensor on the object based on feedback from *Output* and the latest movement information from the agent. *What* predicts object features

