31.1 A 65nm 8.79TOPS/W 23.82mW Mixed-Signal Oscillator-Based NeuroSLAM Accelerator for Applications in Edge Robotics

Jong-Hyeok Yoon, Arijit Raychowdhury

Georgia Institute of Technology, Atlanta, GA

Simultaneous localization and mapping (SLAM) is a quintessential problem in cyber-physical systems with wide-spread applications in mobile robotics, self-driving vehicles, AR, VR, etc. While computational methods [1] and hardware demonstrations [2-5] based on filtering or keyframe techniques are popular, we recognize that ultra-low-power edge-robotics requires circuit solutions that will significantly reduce the power consumption. Interestingly, biological systems can solve SLAM with extreme energy-efficiencies by employing methods that are robust, flexible, and well-integrated into the creatures' sensory systems. Particularly, rodents have shown an extraordinary ability to store and organize visual cues so that a brief sequence of visual cues can globally re-localize the animal. Further, the neural recordings of the rodent hippocampus have led to the discovery of place cells and head direction cells which show striking correlation with mapping tasks. This led to the recent development of a neuromorphic vision-based SLAM algorithm with great success on benchmark tasks [6]. In this paper, we present NeuroSLAM, a spiking neural network (SNN)-based mixed-signal oscillatory circuit, coupled with a lightweight vision system that provides odometry and appearance information. We demonstrate a 65nm test-chip integrated on a mobile robot, performing visual SLAM at 17.27mW (23.82mW) with a net energy-efficiency of 7.25TOPS/W (8.79TOPS/W).

Place cells (X, Y) and head direction cells (θ), which are arranged in a grid, encode spatial location in rodents. Visual odometry allows path integration as the creature moves (self-motion), and the corresponding place and head direction cells fire (Fig. 31.1.1). However, this accumulates error and finally, as the rodent recognizes a previously visited visual cue, neuronal dynamics allows competition to resolve the difference between self-motion and visual cues enabling the rodent to re-localize and construct a robust map. Correspondingly, NeuroSLAM features a 2D mixed-signal SNN-based pose-cell array (X, Y) coupled with digital head-direction (θ) computation in O(N) complexity. Fig. 31.1.1 illustrates the chip-architecture with (1) a vision front-end based on scan-line intensity profiles, (2) visual template (VT) matching, (3) pose-cell control and path integration (odometry) (4) 2 banks of 18.94kB/bank of SRAM for VT storage, and (5) a 7×7 SNN-based circularly connected pose-cell array mimicking the attractor properties of the continuous neural network.

The vision system relies on scan-line intensity profiles [6] where pixel-data are column-wise added, quantized to 4b and 1D-max-pooled (Fig. 31.1.2). Visual odometry requires calculation of: (1) translational velocity (v) based on the difference between the current input and the previous input, and (2) rotational velocity (ω) estimated via shifting the index for the minimum difference [6]. Path integration is conducted by digital integration of ω to obtain θ and virtual pose-cell shift for (X,Y). Loop closure in the SLAM is tracked via VT matching, where a new image needs to be compared to every stored VT. When a new VT is generated, the SRAM stores the input VT and the address corresponding to the pose-cell with maximum energy. Power consumption and latency are minimized during VT matching via: (1) *Dual Thresholds (DT):* If the difference between the image and a stored VT (ΔVT) < TH_{LOW} (lower threshold), the VT is immediately returned reducing further memory access. Only if ΔVT > TH_{HIGH} (higher threshold) for every VT, then a new VT is appended, thus reducing the total memory usage compared to employing a single threshold TH_{LOW}; and if $TH_{LOW} < \Delta VT < TH_{HIGH}$ for every VT, the best matched VT is returned with a full scan of stored VTs, and (2) *Dynamic Indexing (DI):* We exploit the fact that once the agent sees a previously seen visual cue (i.e., VT match at index [j]), the probability of a VT match for the next input is high near [j], by starting the VT search at [j]. DT and DI enable the front-end to reduce the number of memory accesses to 63% of the baseline (Fig. 31.1.2).

Localization of (X,Y) is performed through a bio-mimetic 7×7 SNN-based pose-cell array that implements the dynamics of a neural attractor network. Each pose-cell in the array has excitatory and inhibitory connections to its neighbors with distance-dependent weights (Fig. 31.1.3) [6] and circular boundary conditions to enable continuous tracking while preventing the map-size from being limited. Each pose-cell features: (1) a 5-stage ring-VCO to implement rate-coded spiking neurons, (2) a 4b current DAC (IDAC) for each excitatory (sourcing current) and inhibitory (sinking current) inputs, including global inhibition and current boost (BST) for self-excitatory feedback, and (3) a 4b asynchronous counter-based energy detector that encodes instantaneous pose-cell energy. When a VT match occurs, the vision front-end injects energy into the corresponding pose-cell via E_{inj}, the coupled attractor dynamics (timing diagram shown in Fig. 31.1.3) resolves competition between visual cues and self-motion, and loop closure occurs. SLAM generates the corresponding experience map (Fig. 31.1.4) where error correction during loop closure leads to a redistribution of the distance error across the entire loop. NeuroSLAM allows the agent to continuously move and acquire data and the map is seamlessly updated, corrected and appended. The measured IDAC control voltage (Fig. 31.1.4) is tuned to provide a robust attractor (>500MHz of pose-cell firing rate), while minimizing spurious spiking by non-excited pose-cells and reducing the power consumption. DT and DI reduce the measured latency of the visual system to >65% over benchmark maps and can support >100fps (currently limited by the speed of the camera interface). Each pose-cell can be stalled in its current state and the corresponding energy (frequency) can be individually read out to provide unique observability into SNN attractor dynamics. Fig. 31.1.4 illustrates one instance of the measured competition of the 7×7 array where $(X,Y)_{max}$ is initially at (3,3). After a VT match, a new visual cue injects energy at (6,0) and the neuronal dynamics eventually resolves and corrects the error in odometry.

Figure 31.1.5 illustrates the measured power-performance characteristics and shows an F_{MAX} of 130.8MHz at 23.82mW. This corresponds to a measured peak energy-efficiency of 8.79TOPS/W and 0.203pJ/MAC (Both include power of SRAM, digital and mixed-signal blocks) for the 4b SNN-based pose-cell array. The breakdown of the system power across various building blocks is shown in Fig. 31.1.5 which shows that the pose-cell array, more specifically the IDAC, consumes a significant portion of the system power owing to its continuous-time dynamics. Mismatches and random process variation among the pose-cell VCOs can cause certain low-V_{TH} pose-cells to fire spuriously even when they are weakly excited. In the worst case, a spurious pose-cell can fire faster than the pose-cell with the maximum excitation and can cause algorithmic errors. We quantify the robustness of the system as the pose-cell frequency margin (= frequency of pose-cell with maximum excitation – worst-case spurious firing) and show measurements across multiple dies (Fig. 31.1.5) illustrating >200MHz of margin and correct operation across the entire operating range.

The test-chip is integrated on a mobile robot with an interface to a Raspberry-PI and an embedded camera. We test the system across various standard indoor benchmark arenas and show correct SLAM operation (overlaid on the blueprint of arena 1) and successful loop closure. Fig. 31.1.6 also illustrates how the number of VTs increases with the map-size (i.e. number of frames) for three template indoor arenas. The benchmarking table shows competitive figures-of-merit, ultra-low power (17.27mW) operation and successful system integration and deployment. The die-shot and the chip-characteristics are shown in Fig. 31.1.7.

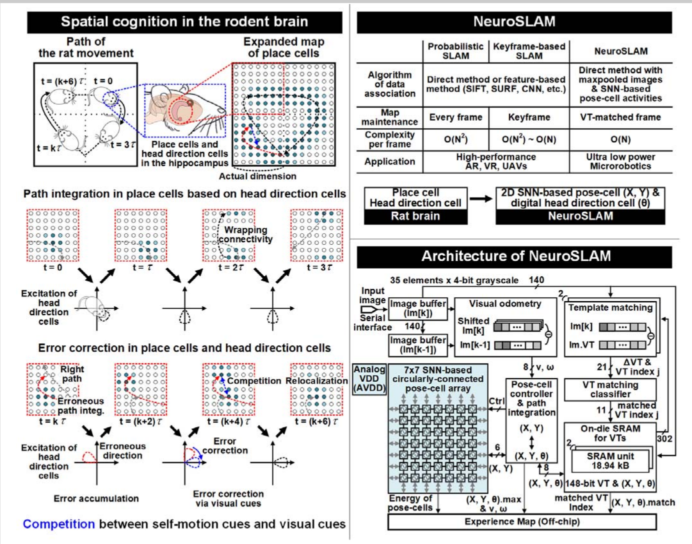
Acknowledgments:
This work was supported by the Semiconductor Research Corporation under the Center for Brain Inspired Computing (C-BRIC) and the National Science Foundation under grant 1640081.

References:

[1] G. Younes et al., “Keyframe-Based Monocular SLAM: Design, Survey, and Future Directions,” *Robotics and Autonomous Systems*, vol. 98, pp. 67-88, 2017.
[2] J. Yoon et al., “A Unified Graphics and Vision Processor With a 0.89 μW/fps Pose Estimation Engine for Augmented Reality,” *IEEE TVLSI*, vol. 21, no. 2, pp. 206-216, 2013.
[3] I. Hong et al., “A 27 mW Reconfigurable Marker-Less Logarithmic Camera Pose Estimation Engine for Mobile Augmented Reality Processor,” *IEEE J. of Solid-State Circuits*, vol. 50, no. 11, pp. 2513-2523, Nov. 2015.
[4] A. Suleiman et al., “Navion: A Fully Integrated Energy-Efficient Visual-Inertial Odometry Accelerator for Autonomous Navigation of Nano Drones,” *IEEE Symp. VLSI Circuits*, pp. 133-134, 2018.
[5] Z. Li et al., “An 879GOPS 243mW 80fps VGA Fully Visual CNN-SLAM Processor for Wide-Range Autonomous Exploration,” *ISSCC*, pp. 134-136, Feb. 2019.
[6] M. J. Milford et al., “Mapping a Suburb With a Single Camera Using a Biologically Inspired SLAM System,” *IEEE Transactions on Robotics*, vol. 24, no. 5, pp. 1038-1053, Oct. 2008.

978-1-7281-3205-1/20/$31.00 ©2020 IEEE

Authorized licensed use limited to: Georgia Institute of Technology. Downloaded on June 28,2021 at 12:56:21 UTC from IEEE Xplore. Restrictions apply.


Spatial cognition in the rodent brain
Path of the rat movement
Expanded map of place cells
Place cells and head direction cells in the hippocampus
Actual dimension
Path integration in place cells based on head direction cells
Wrapping connectivity
Excitation of head direction cells
Error correction in place cells and head direction cells
Right path
Erroneous path integ.
Competition
Relocalization
Erroneous direction
Error correction
Error accumulation
Error correction via visual cues
Competition between self-motion cues and visual cues
NeuroSLAM
Architecture of NeuroSLAM
35 elements x 4-bit grayscale
Input image
Serial interface
Image buffer (Im[k])
Image buffer (Im[k-1])
Visual odometry
Template matching
Analog VDD (AVDD)
7x7 SNN-based circularly-connected pose-cell array
Pose-cell controller & path integration
VT matching classifier
On-die SRAM for VTs
SRAM unit 18.94 kB
148-bit VT & (X, Y, θ)
Energy of pose-cells
matched VT index
Experience Map (Off-chip)

	Probabilistic SLAM	Keyframe-based SLAM	NeuroSLAM
Algorithm of data association	Direct method or feature-based method (SIFT, SURF, CNN, etc.)		Direct method with maxpooled images & SNN-based pose-cell activities
Map maintenance	Every frame	Keyframe	VT-matched frame
Complexity per frame	$O(N^2)$	$O(N^2) \sim O(N)$	$O(N)$
Application	High-performance AR, VR, UAVs		Ultra low power Microrobotics

Place cell
Head direction cell — Rat brain → 2D SNN-based pose-cell (X, Y) & digital head direction cell (θ) — NeuroSLAM

Figure 31.1.1: Motivation for NeuroSLAM, comparison with other algorithms and the overall system architecture illustrating the key design components.

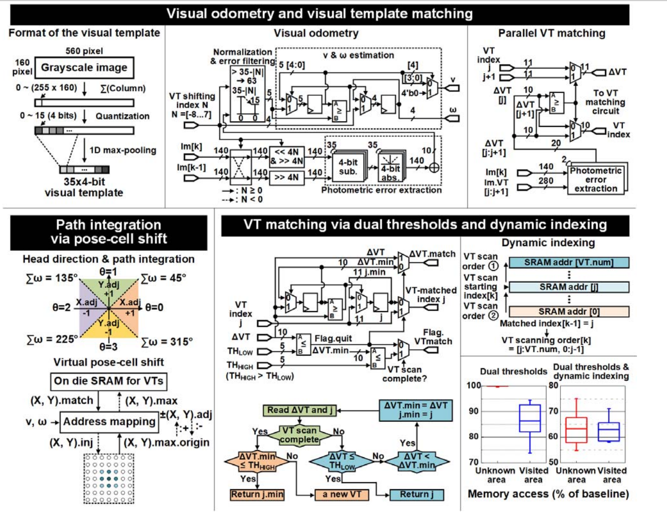


Figure 31.1.2: Proposed scan-line intensity profile-based visual odometry and template matching, and path integration via virtual pose-cell shift in the pose-cell array.

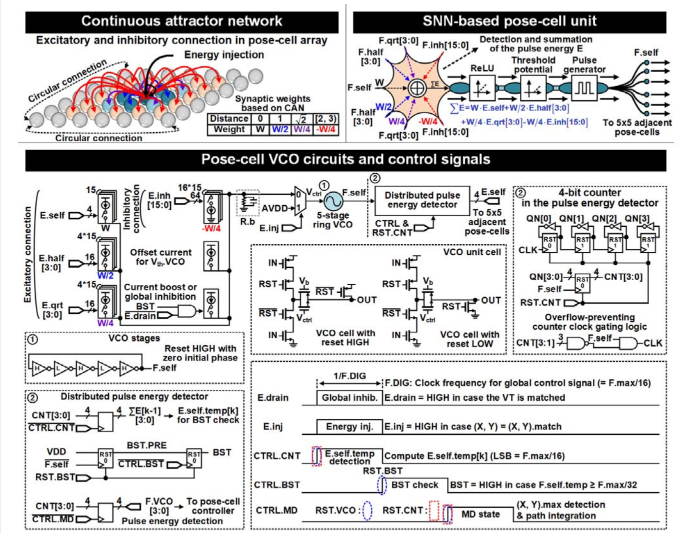

Continuous attractor network
Excitatory and inhibitory connection in pose-cell array
Energy injection
Circular connection
Synaptic weights based on CAN
Distance 0 1 √2 [2, 3]
Weight W W/2 W/4 -W/4
SNN-based pose-cell unit
F.qrt[3:0]
F.inh[15:0]
Detection and summation of the pulse energy E
Threshold potential
Pulse generator
ReLU
F.self
F.half
To 5x5 adjacent pose-cells
ΣE=W·E.self+W/2·E.half[3:0] +W/4·E.qrt[3:0]-W/4·E.inh[15:0]
Pose-cell VCO circuits and control signals
Excitatory connection
Inhibitory connection
Offset current for Vth,VCO
Current boost or global inhibition
5-stage ring VCO
Distributed pulse energy detector
4-bit counter in the pulse energy detector
VCO unit cell
VCO cell with reset HIGH
VCO cell with reset LOW
Overflow-preventing counter clock gating logic
VCO stages
Reset HIGH with zero initial phase
Distributed pulse energy detector
Pulse energy detection
1/F.DIG
F.DIG: Clock frequency for global control signal (= F.max/16)
E.drain = HIGH in case the VT is matched
E.inj = HIGH in case (X, Y) = (X, Y).match
Compute E.self.temp[k] (LSB = F.max/16)
BST = HIGH in case F.self.temp ≥ F.max/32
(X, Y).max detection & path integration

Figure 31.1.3: Proposed oscillator-based mixed-signal pose-cell design, circuit components and timing diagrams.

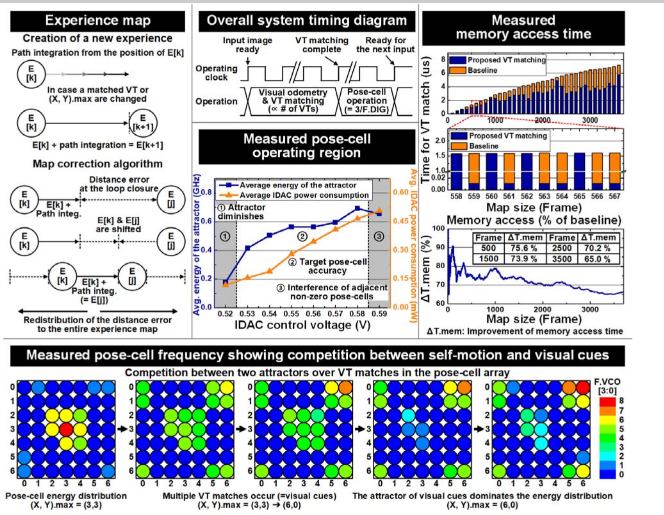


Figure 31.1.4: Experience map management, overall system timing diagram, and memory access in proposed visual template matching, pose-cell operating region, and continuous attractor network-based pose-cell activities.

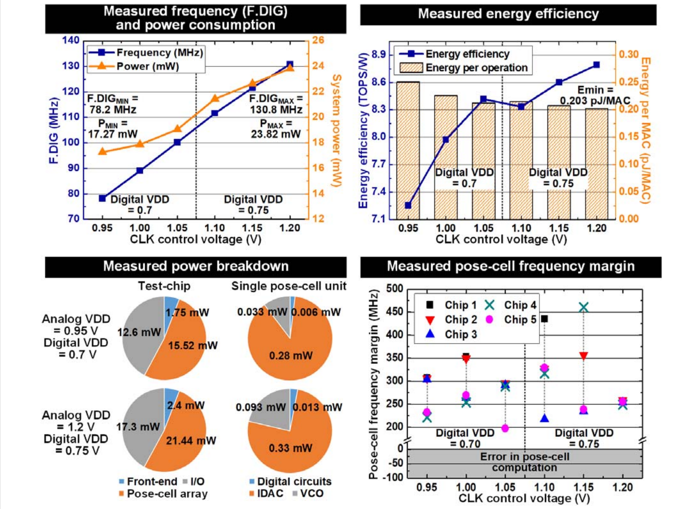
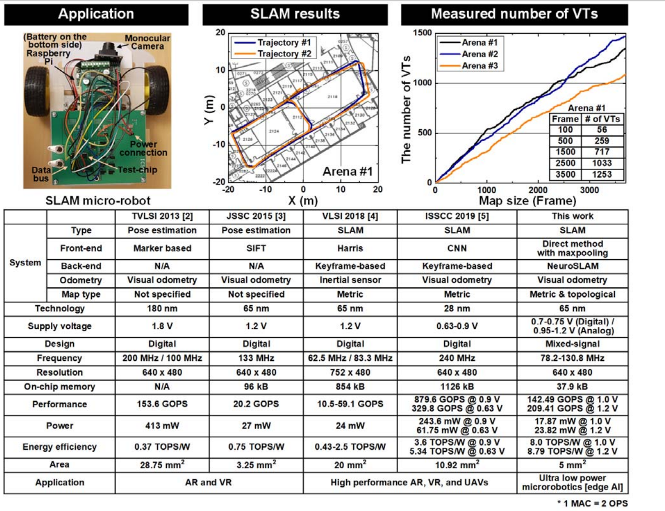


Figure 31.1.5: Power consumption, operating frequencies, energy efficiency, power breakdown across various design components, and resiliency of the system showing pose-cell frequency margin across multiple test-chips.

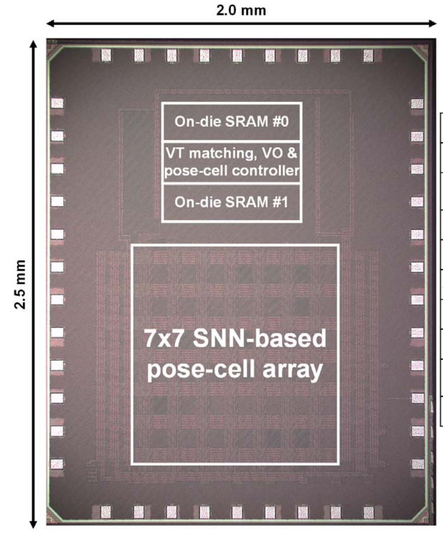


		TVLSI 2013 [2]	JSSC 2015 [3]	VLSI 2018 [4]	ISSCC 2019 [5]	This work
System	Type	Pose estimation	Pose estimation	SLAM	SLAM	SLAM
	Front-end	Marker based	SIFT	Harris	CNN	Direct method with maxpooling
	Back-end	N/A	N/A	Keyframe-based	Keyframe-based	NeuroSLAM
	Odometry	Visual odometry	Visual odometry	Inertial sensor	Visual odometry	Visual odometry
	Map type	Not specified	Not specified	Metric	Metric	Metric & topological
Technology		180 nm	65 nm	65 nm	28 nm	65 nm
Supply voltage		1.8 V	1.2 V	1.2 V	0.63-0.9 V	0.7-0.75 V (Digital) / 0.95-1.2 V (Analog)
Design		Digital	Digital	Digital	Digital	Mixed-signal
Frequency		200 MHz / 100 MHz	133 MHz	62.5 MHz / 83.3 MHz	240 MHz	78.2-130.8 MHz
Resolution		640 x 480	640 x 480	752 x 480	640 x 480	640 x 480
On-chip memory		N/A	96 kB	854 kB	1126 kB	37.9 kB
Performance		153.6 GOPS	20.2 GOPS	10.5-59.1 GOPS	879.6 GOPS @ 0.9 V 329.8 GOPS @ 0.63 V	142.49 GOPS @ 1.0 V 209.41 GOPS @ 1.2 V
Power		413 mW	27 mW	24 mW	243.6 mW @ 0.9 V 61.75 mW @ 0.63 V	17.87 mW @ 1.0 V 23.82 mW @ 1.2 V
Energy efficiency		0.37 TOPS/W	0.75 TOPS/W	0.43-2.5 TOPS/W	3.6 TOPS/W @ 0.9 V 5.34 TOPS/W @ 0.63 V	8.0 TOPS/W @ 1.0 V 8.79 TOPS/W @ 1.2 V
Area		28.75 mm^2	3.25 mm^2	20 mm^2	10.92 mm^2	5 mm^2
Application		AR and VR		High performance AR, VR, and UAVs		Ultra low power microrobotics [edge AI]

* 1 MAC = 2 OPS

Figure 31.1.6: Application of the test-chip to SLAM in mobile micro-robotics, measurement results of SLAM operation across benchmark arenas, and comparison with competing designs.

Authorized licensed use limited to: Georgia Institute of Technology. Downloaded on June 28,2021 at 12:56:21 UTC from IEEE Xplore. Restrictions apply.


2.0 mm
2.5 mm
On-die SRAM #0
VT matching, VO & pose-cell controller
On-die SRAM #1
7x7 SNN-based pose-cell array

Technology	65 nm 1P9M CMOS
Die area	2.0 mm x 2.5 mm
On-chip memory	37.9 kB
Frequency	78.22-130.8 MHz
Digital VDD	0.7-0.75 V
Analog VDD	0.95-1.2 V
I/O VDD	2.5 V
Power	17.27-23.82 mW
Energy efficiency	7.25-8.79 TOPS/W
Package	QFN48

Figure 31.1.7: Microphotograph of the test-chip, chip characteristics and summary of performance.

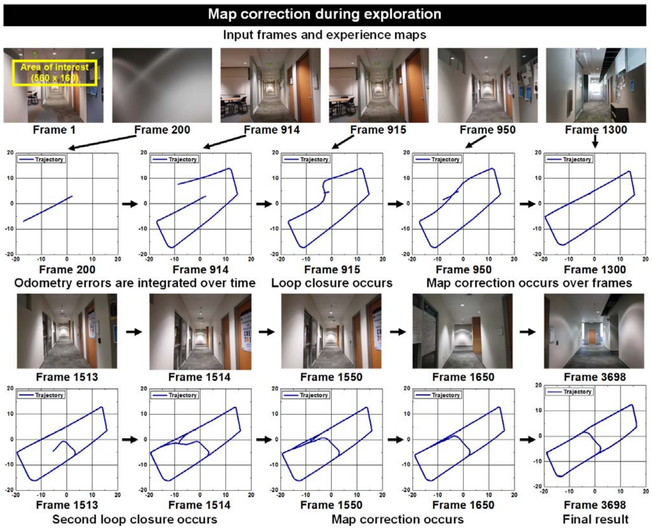


Figure 31.1.S1: Operation of the test-chip in a real environment showing map correction in the experience map during exploration.

Authorized licensed use limited to: Georgia Institute of Technology. Downloaded on June 28,2021 at 12:56:21 UTC from IEEE Xplore. Restrictions apply.