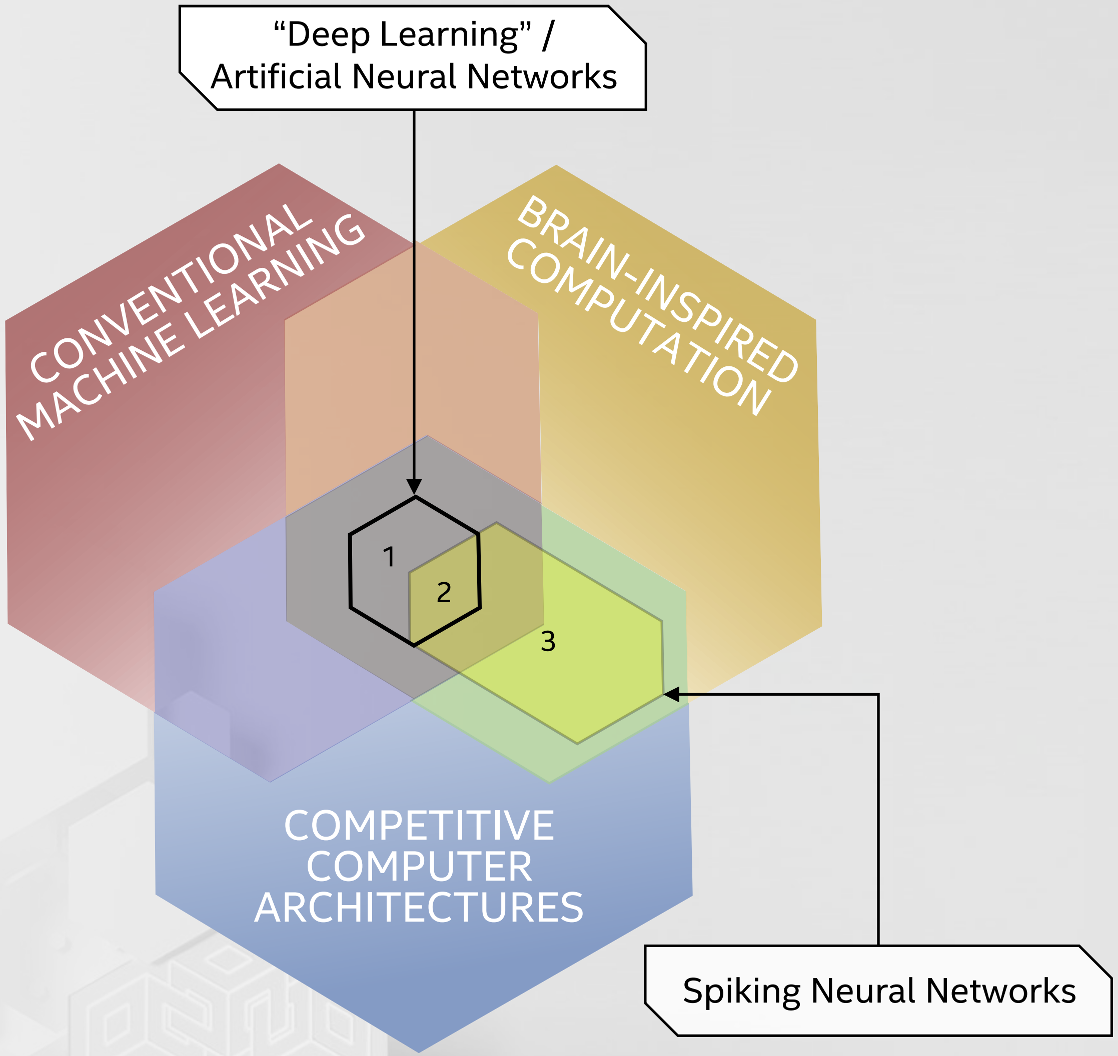
Scaling Up Neuromorphic Systems for Breakthroughs in Computing

Mike Davies | Director, Neuromorphic Computing Lab | Intel Labs

November 18, 2019

Legal Information

Neuromorphic Computing Exploration Space



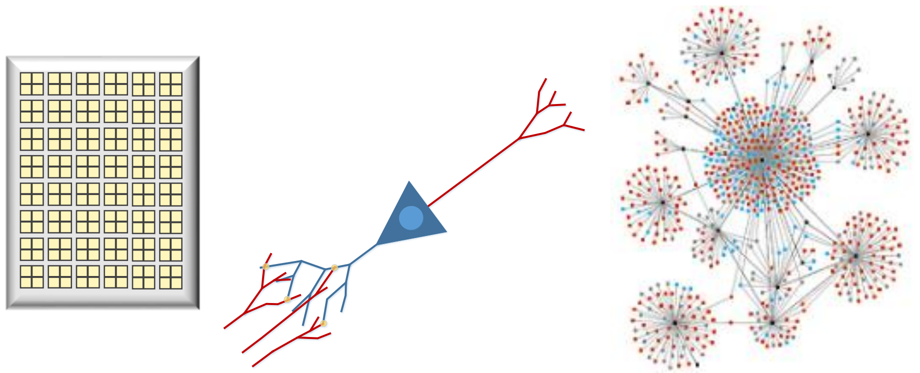
RESEARCH GOALS

- **Broad class** of brain-inspired computation
- **Efficient** hardware implementations
- **Scalable** from small to large problems and systems

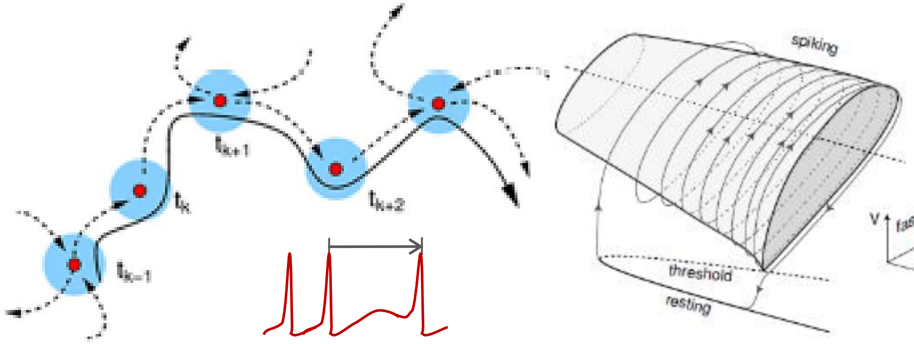
EXAMPLE WORKLOADS

- Learning without cloud assistance
- Learning with sparse supervision
- Online and lifelong learning
- Probabilistic inference and learning
- Sparse coding
- Associative memory, similarity matching
- Nonlinear adaptive control (robotics)
- SLAM and path planning
- Constraint satisfaction
- Dynamical systems modeling

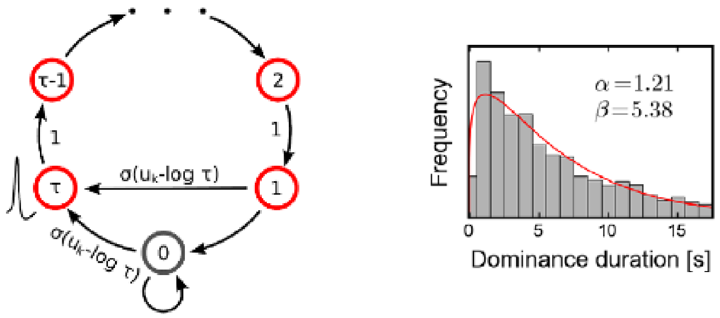
Some Principles of Neural Computation



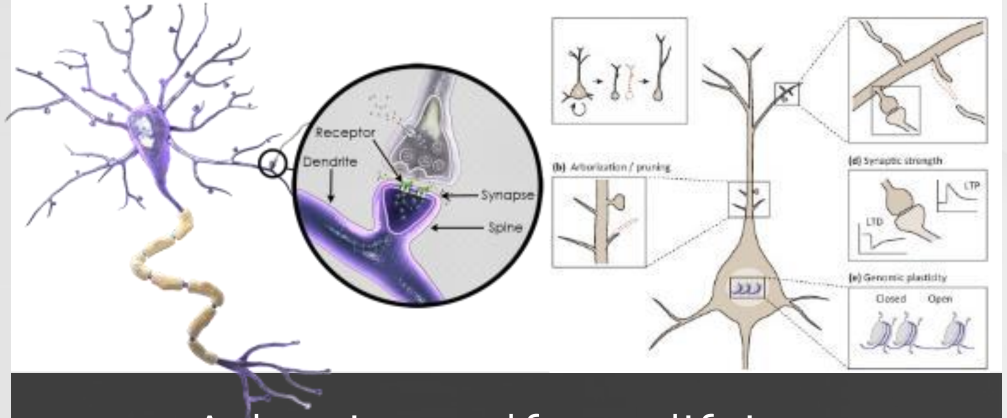
Fine-grained parallelism
with massive fanout



Event-driven computation
with time

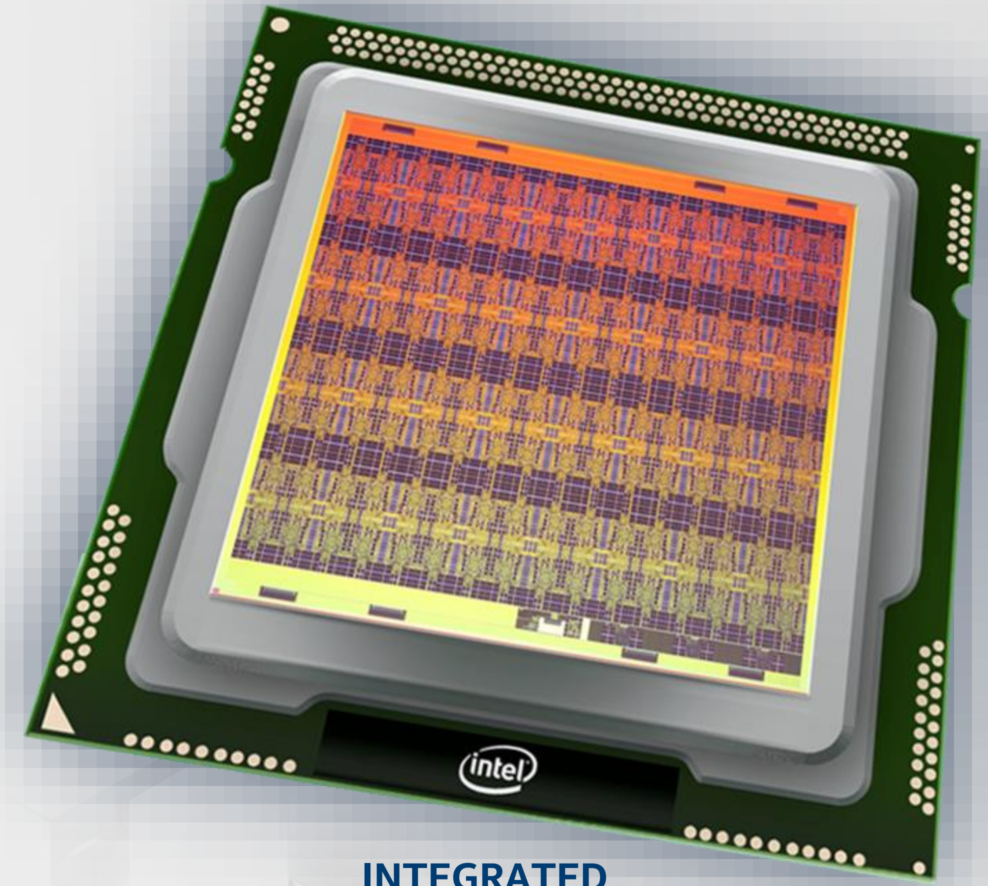


Low precision and stochastic



Adaptive, self-modifying

Our Loihi Research Chip



**INTEGRATED
MEMORY + COMPUTE
NEUROMORPHIC ARCHITECTURE**

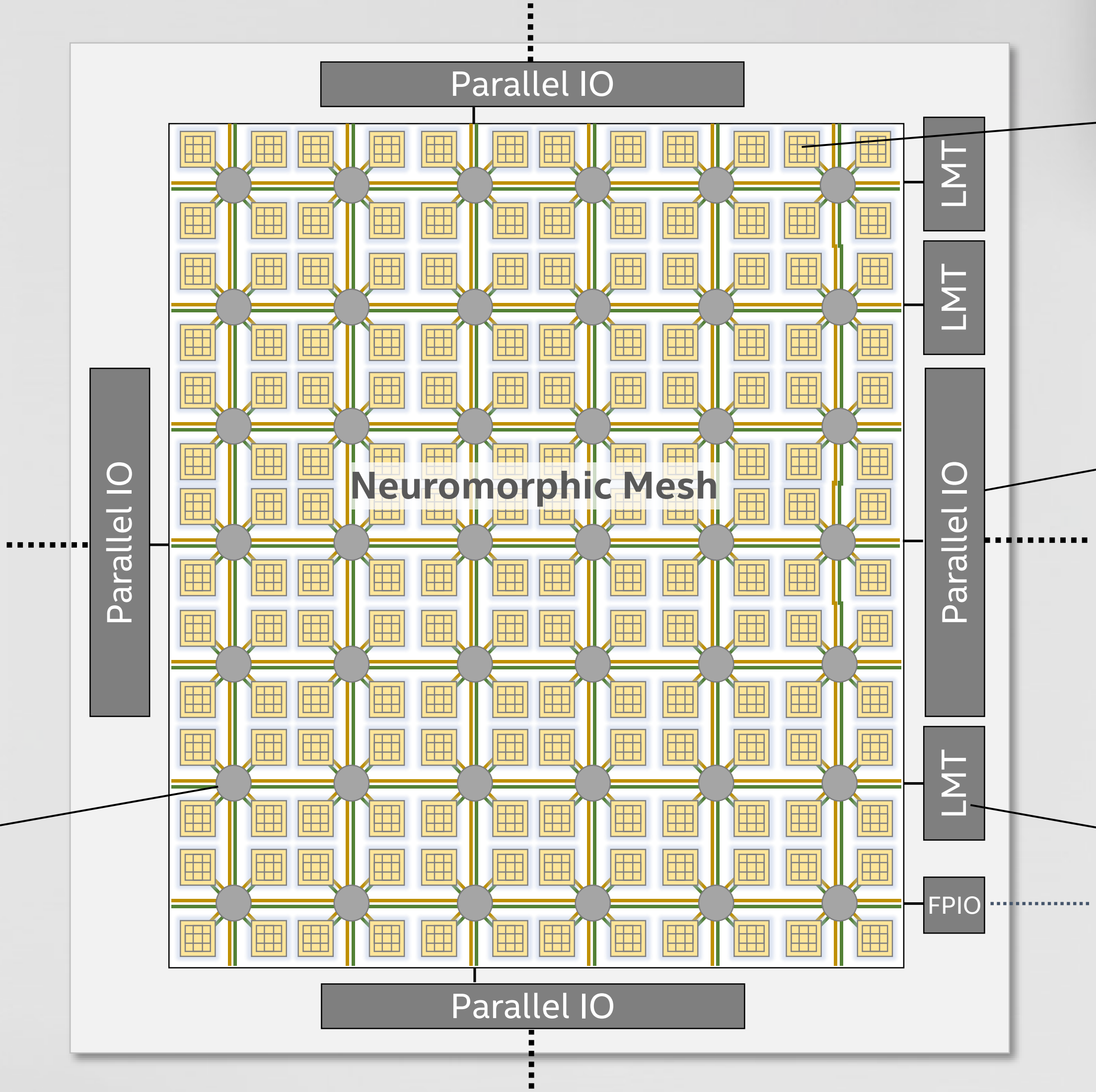
KEY PROPERTIES

- 128 neuromorphic cores supporting up to 128k neurons and 128M synapses with an **advanced spiking neural network feature set.**
- Supports **highly complex neural network topologies**
- **Scalable on-chip learning** capabilities to support an unprecedented range of learning algorithms
- Fully digital **asynchronous** implementation
- Fabricated in Intel's **14nm FinFET process** technology

Davies et al, "Loihi: A Neuromorphic Manycore Processor with On-Chip Learning." IEEE Micro, Jan/Feb 2018.

Chip Architecture

Technology:	14nm
Die Area:	60 mm2
Neuro cores:	128 cores
x86 cores:	3 LMT cores
Max # neurons:	128K neurons
Max # synapses:	128M synapses
Transistors:	2.07 billion
Memory:	33 MB
Efficiency:	42 GOPS/W



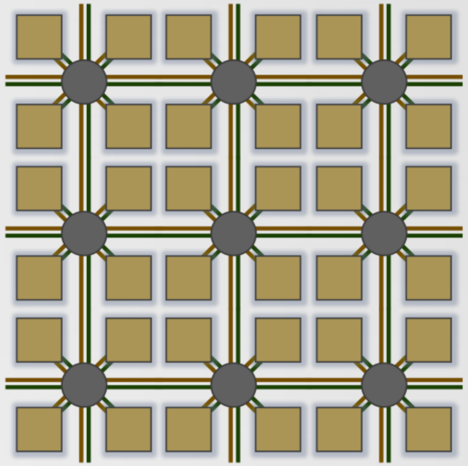
Neuromorphic core
- Leaky integrate-and-fire neuron model
- Programmable learning
- 128 KB synaptic memory
- Up to 1,024 neurons
- Asynchronous design

Parallel off-chip interfaces
- Two-phase asynchronous
- Single-ended signaling
- 100-200 MB/s BW

Low-overhead NoC fabric
- 8x16-core 2D mesh
- Scalable to 1000's cores
- Dimension order routed
- Two physical fabrics
- 8 GB/s per hop

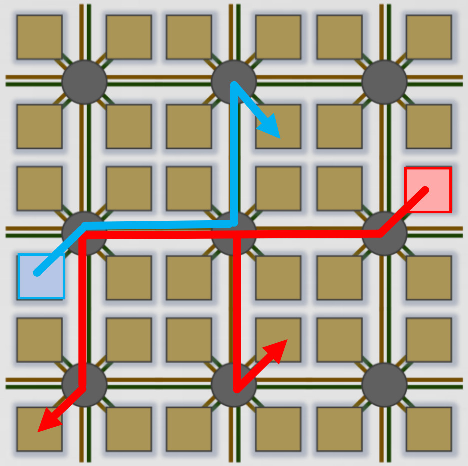
Embedded x86 processors
- Efficient spike-based communication with neuromorphic cores
- Data encoding/decoding
- Network configuration
- Synchronous design

Mesh Operation: Fine-Grained Synchronization



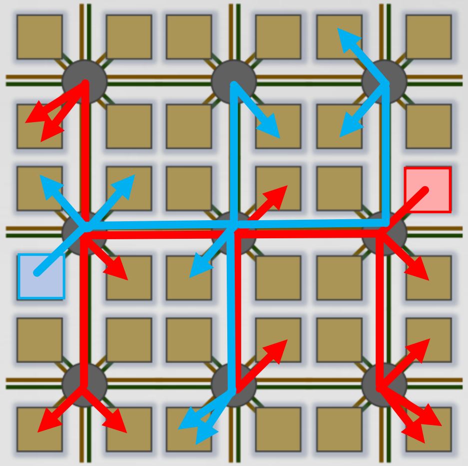
Time step T begins.

Cores update dynamic neuron state and evaluate firing thresholds

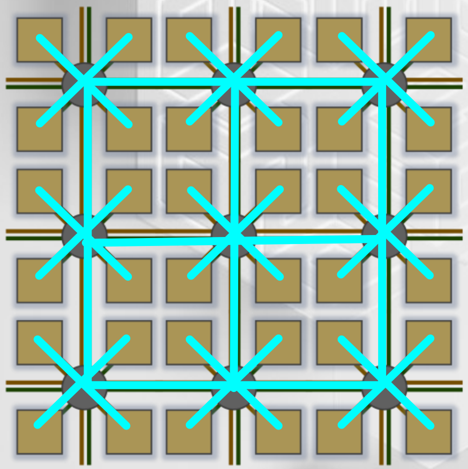


Above-threshold neurons send spike messages to fanout cores

(Two neuron firings shown.)

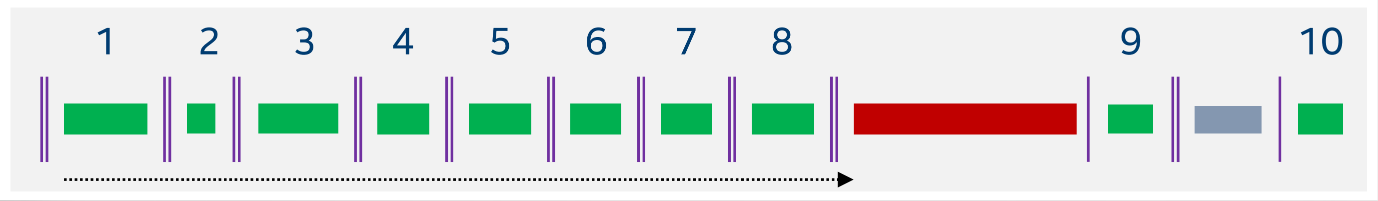


All neurons that fire in time T route their spike messages to all destination cores.

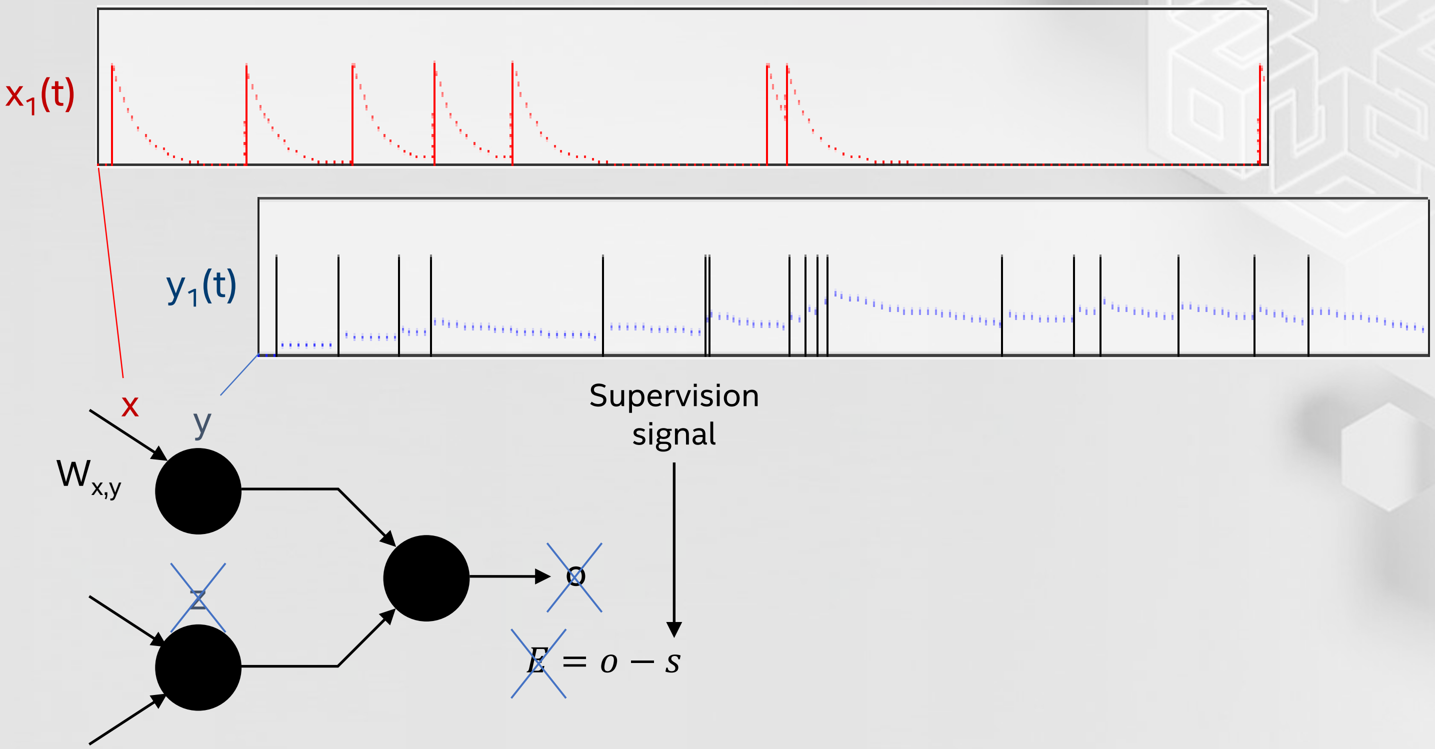


Barrier Synchronization messages exchanged between all cores.

When complete, time advances to time step T+1.



Programmable Learning with Local Plasticity

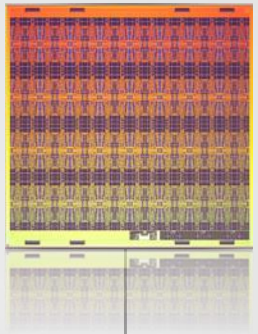


Scalable from Embedded IP to Datacenter

EMBEDDED

Bare Loihi Chip | Kapoho Bay | Wolf Mountain | Nahuku | Pohoiki Beach

DATACENTER



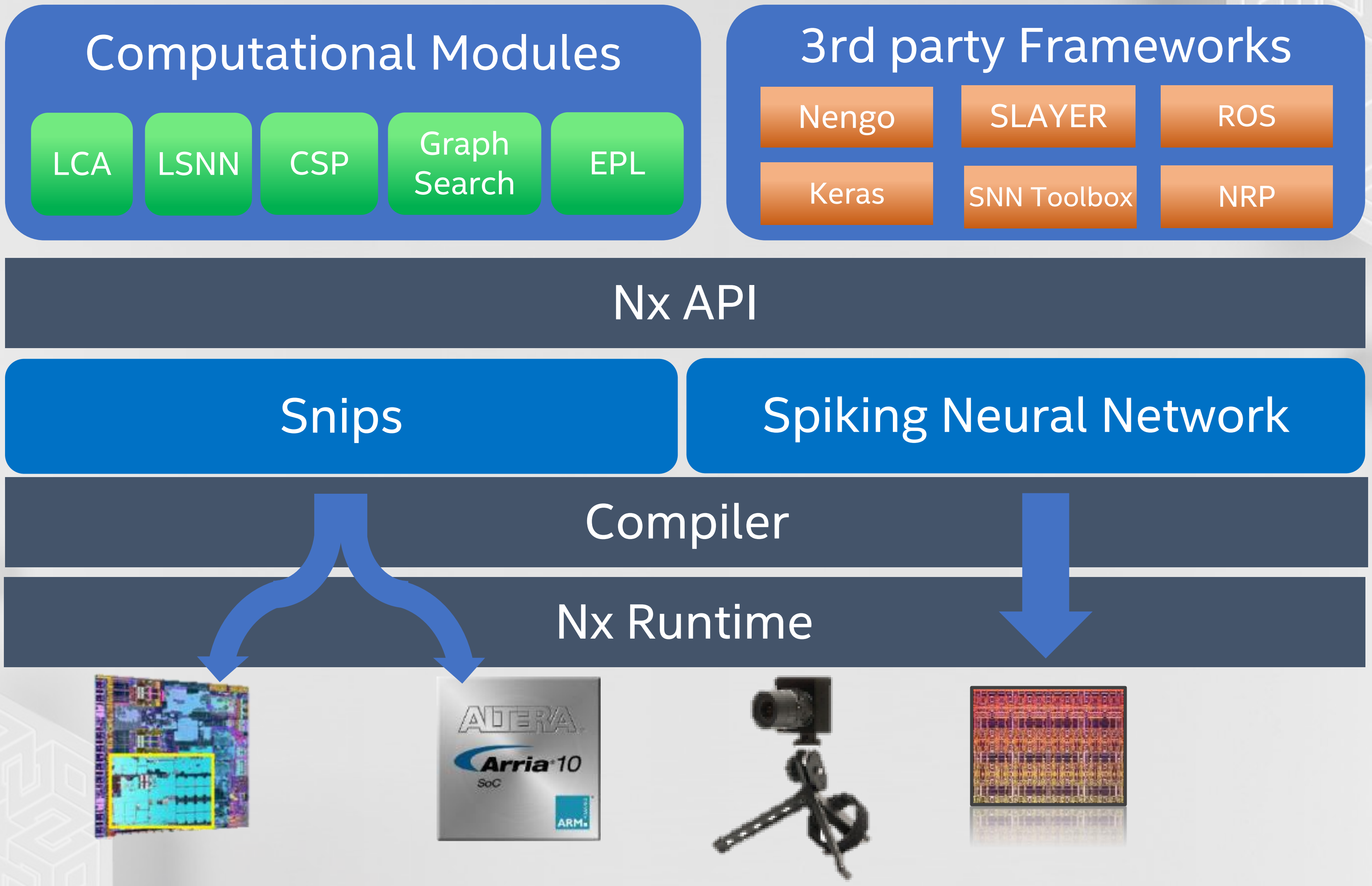
Scalable IP that can be embedded in SoCs (1-100mW)

Single chip form factors for edge devices, e.g. processing event-based camera input (<1W)

Multi-chip form factors for real-time AI, SLAM, planning, optimization

Rack-mounted datacenter appliance for analytics, workload acceleration, virtual robotics modeling, dynamical systems modeling, neuroscience research

Nx SDK Software Architecture



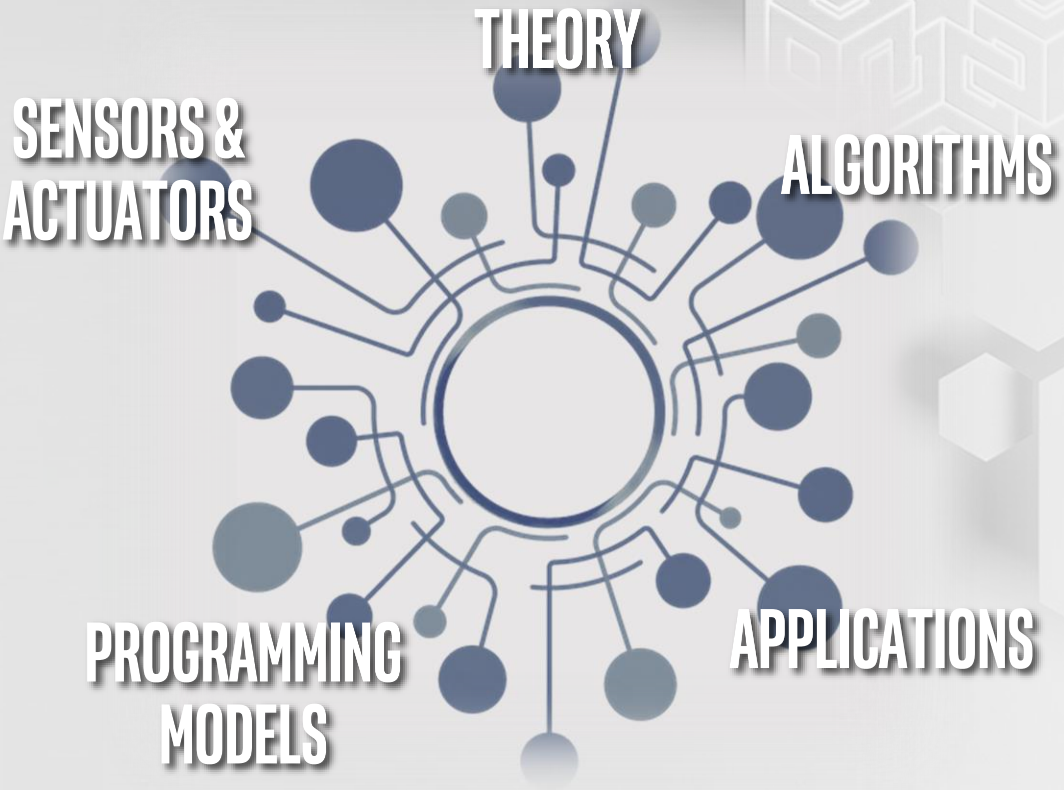
Intel Neuromorphic Research Community

Collaborating to Accelerate the Research


Intel News
@intelnews
Follow

Intel creates neuromorphic research community to advance 'Loihi' test chip: intel.ly/2oxQJT1

3:04 PM - 1 Mar 2018



75 ENGAGED ACADEMIC, GOVERNMENT, AND INDUSTRY GROUPS

Email inrc_interest@intel.com to get involved!

accenture

AIRBUS

ABR
Applied Brain Research

Cardiff University
Prifysgol Caerdydd
Home of Innovation
Cartref Arloesedd

DONDERS INSTITUTE

Drexel University

dsrl dynamical systems & risk laboratory
UCD Dublin

GE

Georg-August-Universität Göttingen

HITACHI
Inspire the Next

IDEAS
Engineering & Technology

King's College London

NUS
National University of Singapore

PennState

RIVERSIDE RESEARCH

RUTGERS

Syracuse University

Technical University of Munich
TUM

имени И.Н.Ульянова
ЧГУ

Berkeley
University of California

UCI University of California, Irvine

University of California
MERCED

UC SANTA BARBARA

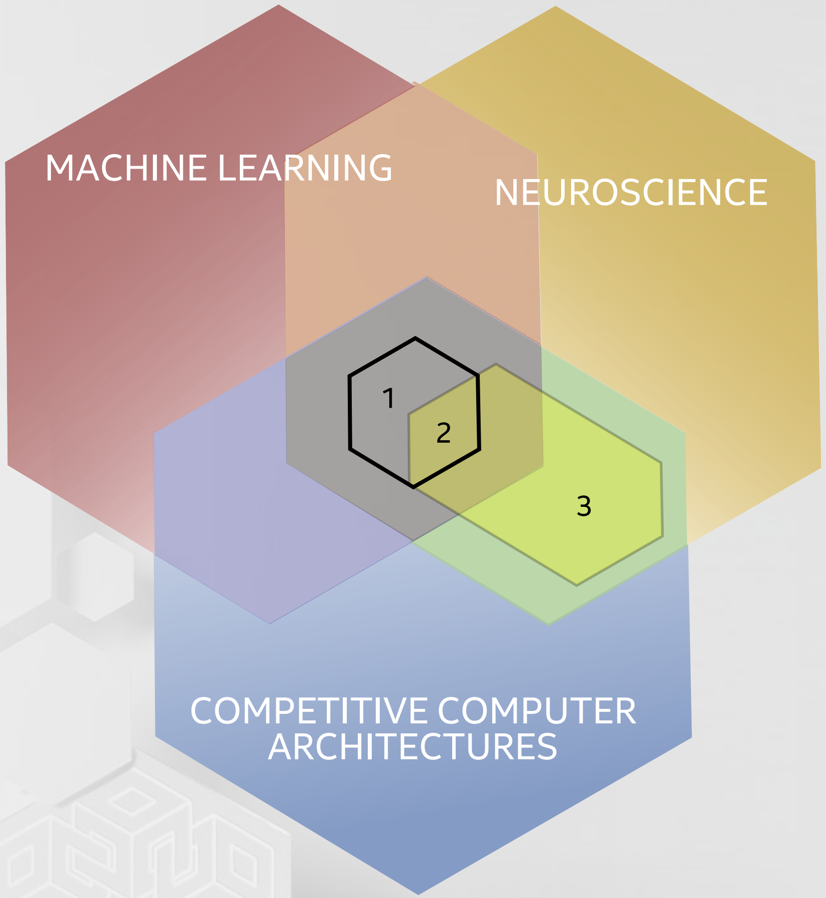
Universitäts Klinikum Freiburg

UEC Tokyo 電気通信大学
The University of Electro-Communications

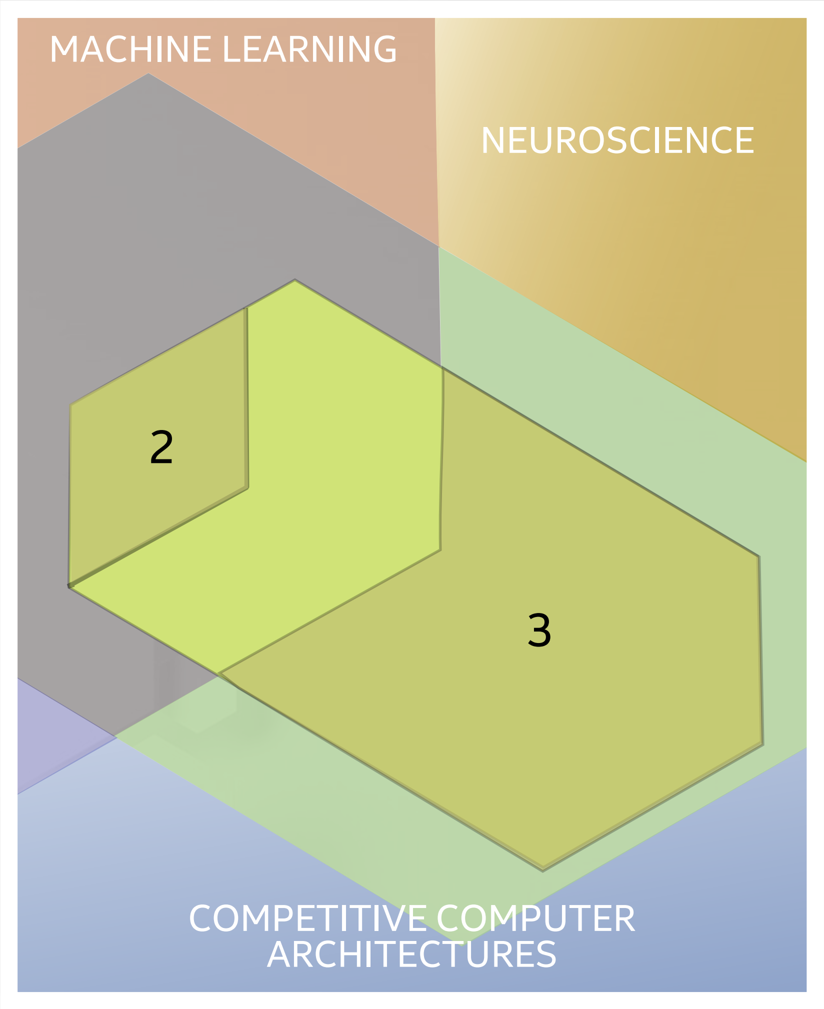
University of Zurich
ETH zürich

Washington State University
Vancouver

The Challenge: SNN Algorithm Discovery

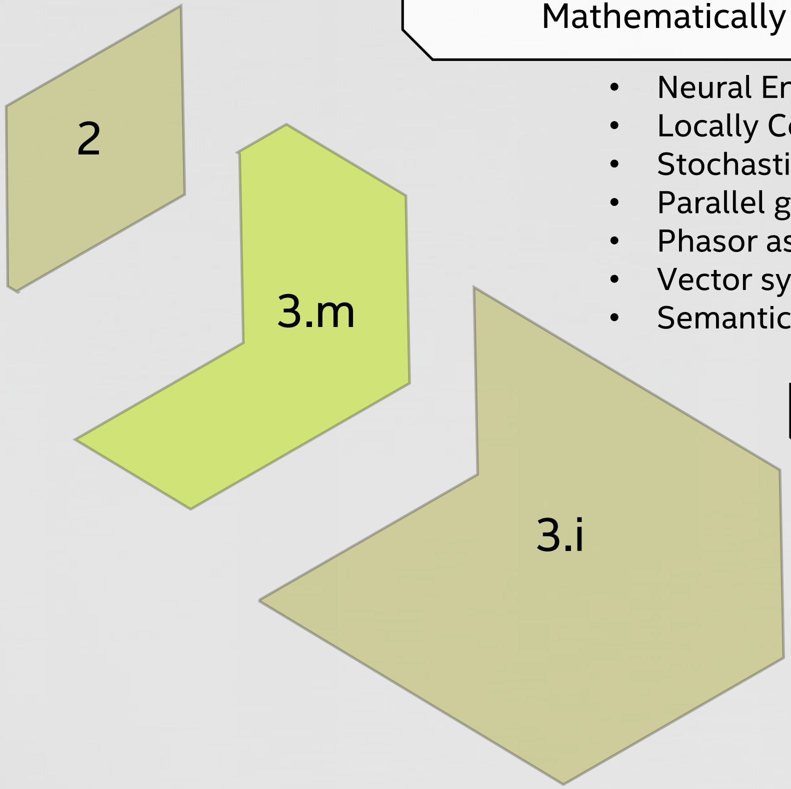


The Challenge: SNN Algorithm Discovery



Deep Learning Derived Approaches

- DNN -> SNN conversion
- SNN backpropagation
- Online SNN pseudo-backprop



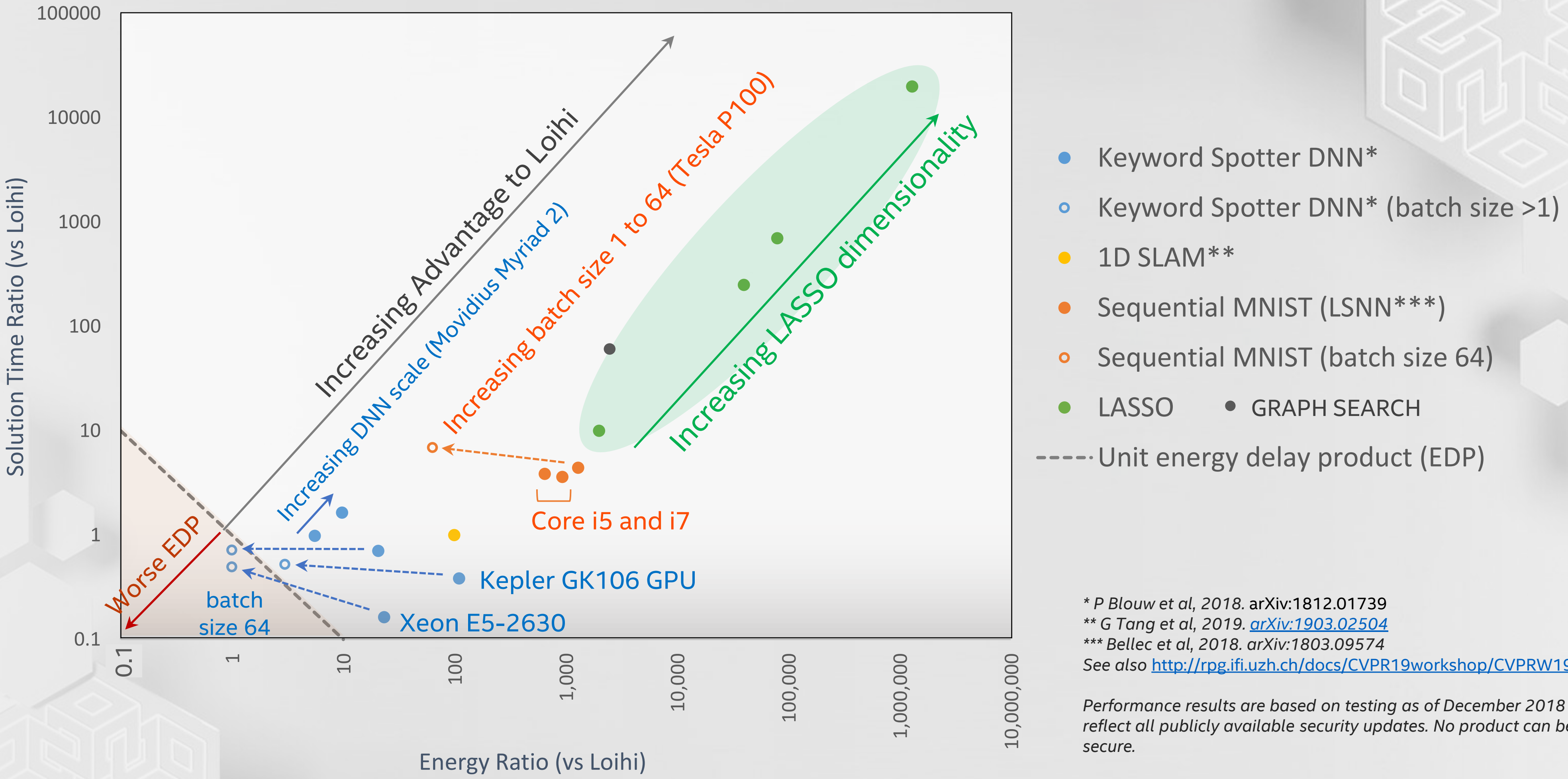
Mathematically Formalized

- Neural Engineering Framework (NEF)
- Locally Competitive Algorithm for LASSO
- Stochastic SNNs for solving CSPs
- Parallel graph search
- Phasor associative memories
- Vector symbolic architectures (VSA)
- Semantic pointer architecture (SPA)

New Ideas Guided by Neuroscience

- Olfaction-inspired rapid learning
- SLAM
- Dynamic Neural Fields
- Evolutionary search
- Cortical models

Loihi Quantitative Results Summary



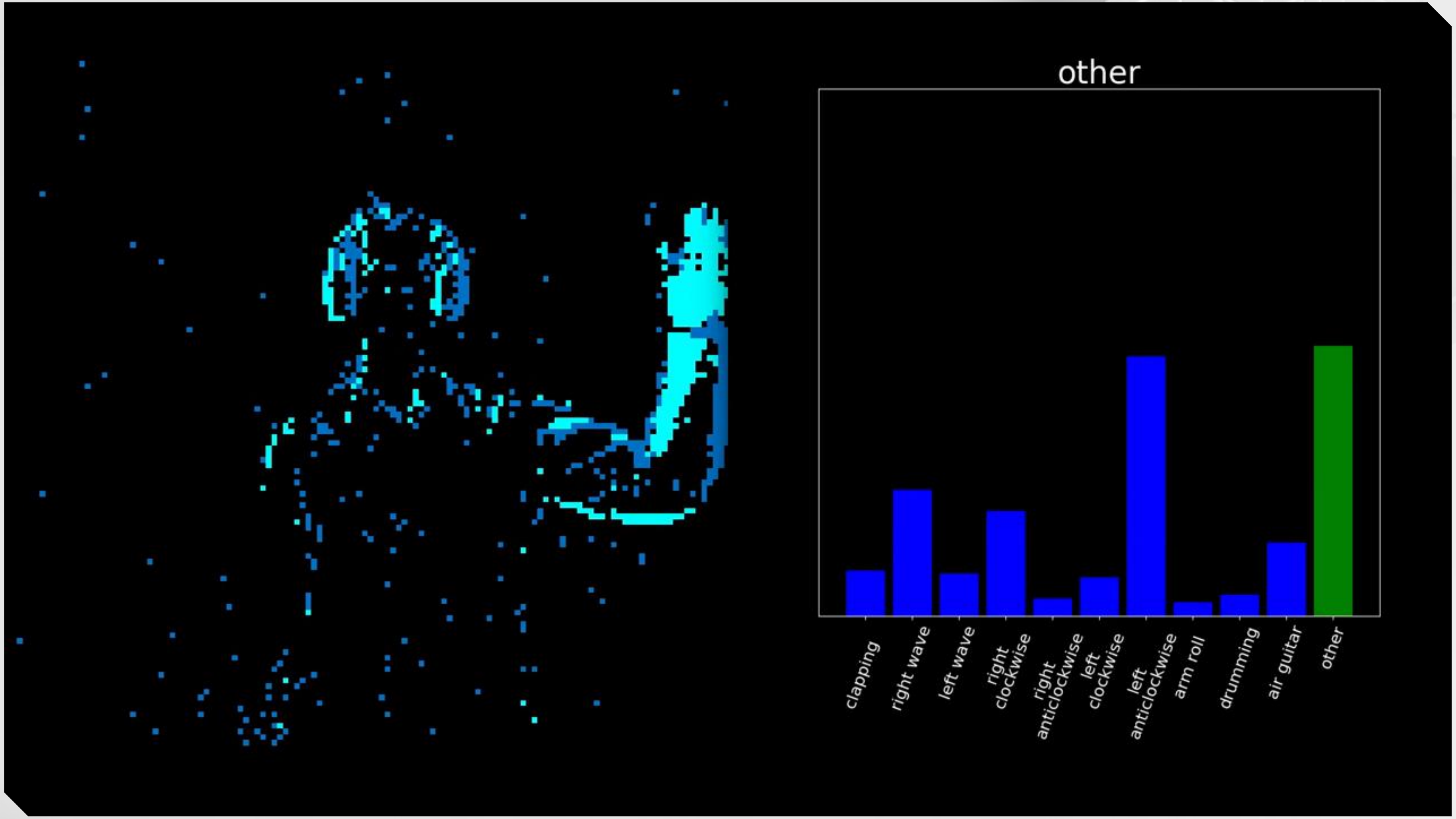
* *P Blouw et al, 2018.* arXiv:1812.01739
** *G Tang et al, 2019. arXiv:1903.02504*
*** *Bellec et al, 2018. arXiv:1803.09574*
See also http://rpg.ifi.uzh.ch/docs/CVPR19workshop/CVPRW19_Mike_Davies.pdf

Performance results are based on testing as of December 2018 and may not reflect all publicly available security updates. No product can be absolutely secure.

Event-Based Camera Gesture Recognition

<u>DAVIS240C</u>*
5mW static
5mW dynamic
1ms latency

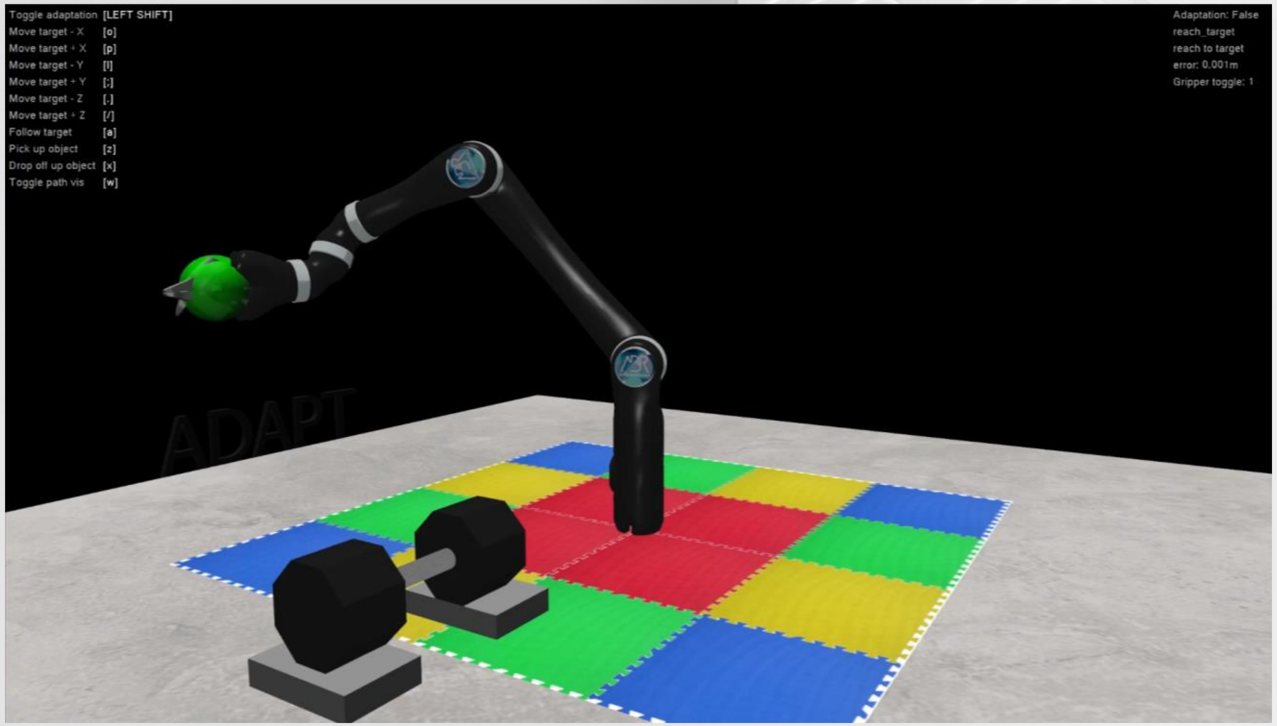
<u>Loihi</u>†
36mW static
7mW dynamic
10ms latency

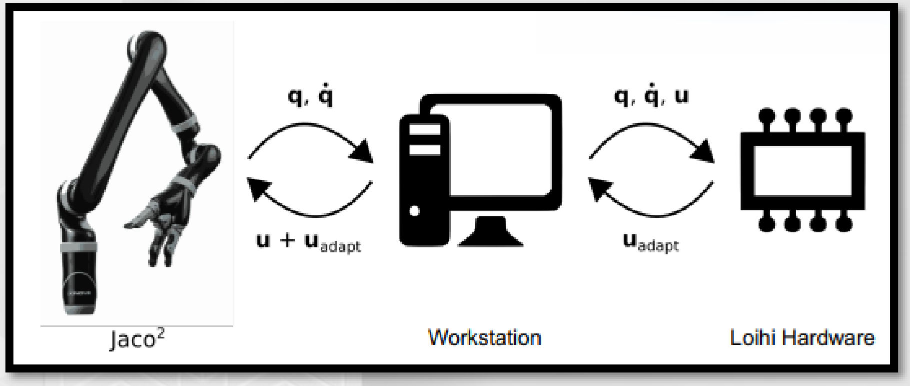


* **iniVation DAVIS 240C** performance numbers obtained from published specifications
† **Intel Loihi** measurements obtained using NxSDK v0.85 running on Wolf Mountain
Performance results are based on testing as of October 2019 and may not reflect all publicly available security updates. No product can be absolutely secure.

Adaptive Control of a Robot Arm Using Loihi

- SNN adaptive dynamic controller implemented on Loihi allows a robot arm to adjust in real time to nonlinear, unpredictable changes in system mechanics[1][2].
- Result outperforms standard PD & PID control algorithms.





[1] DeWolf, T., Stewart, T. C., Slotine, J. J., & Eliasmith, C. (2016, November). A spiking neural model of adaptive arm control. In *Proc. R. Soc. B* (Vol. 283, No. 1843, p. 20162134). The Royal Society.
[2] Eliasmith, "Building applications with next generation neuromorphic hardware." *NICE Workshop 2018*

LASSO Sparse Coding

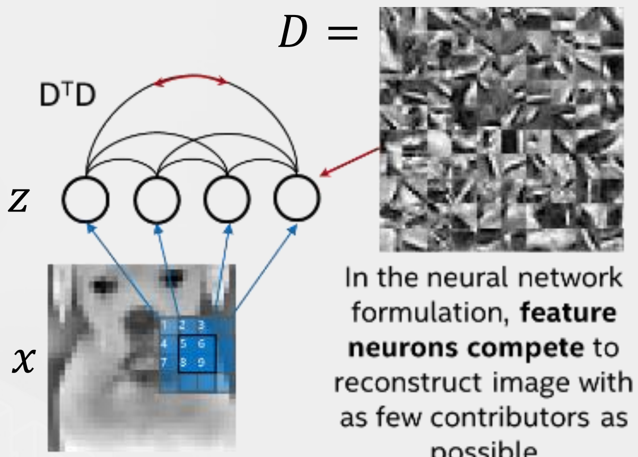
The Spiking Locally Competitive Algorithm (S-LCA)

Problem

$$\min_z \frac{1}{2}\|x - Dz\|_2^2 + \lambda\|z\|_1$$

Input | Reconstruction | Sparse regularization

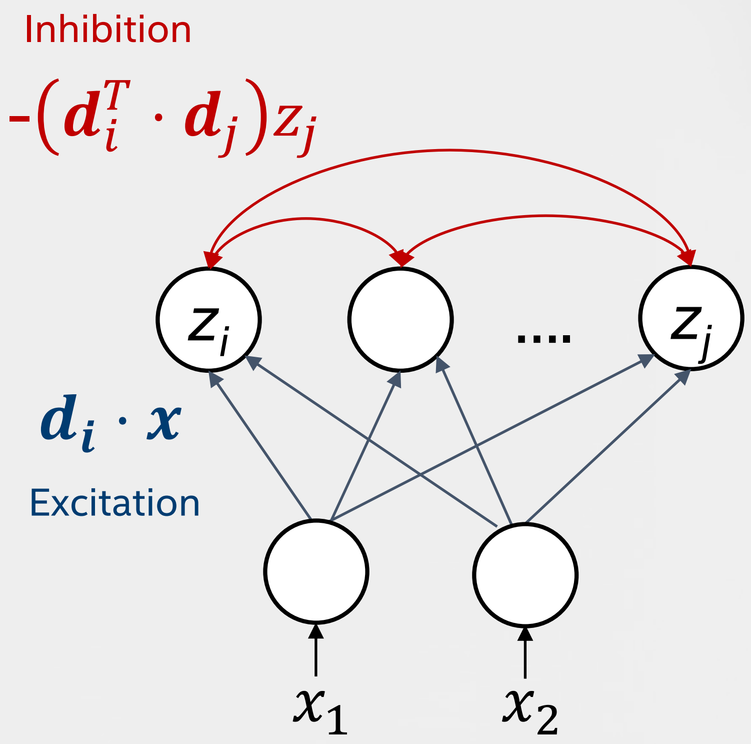
Implementation



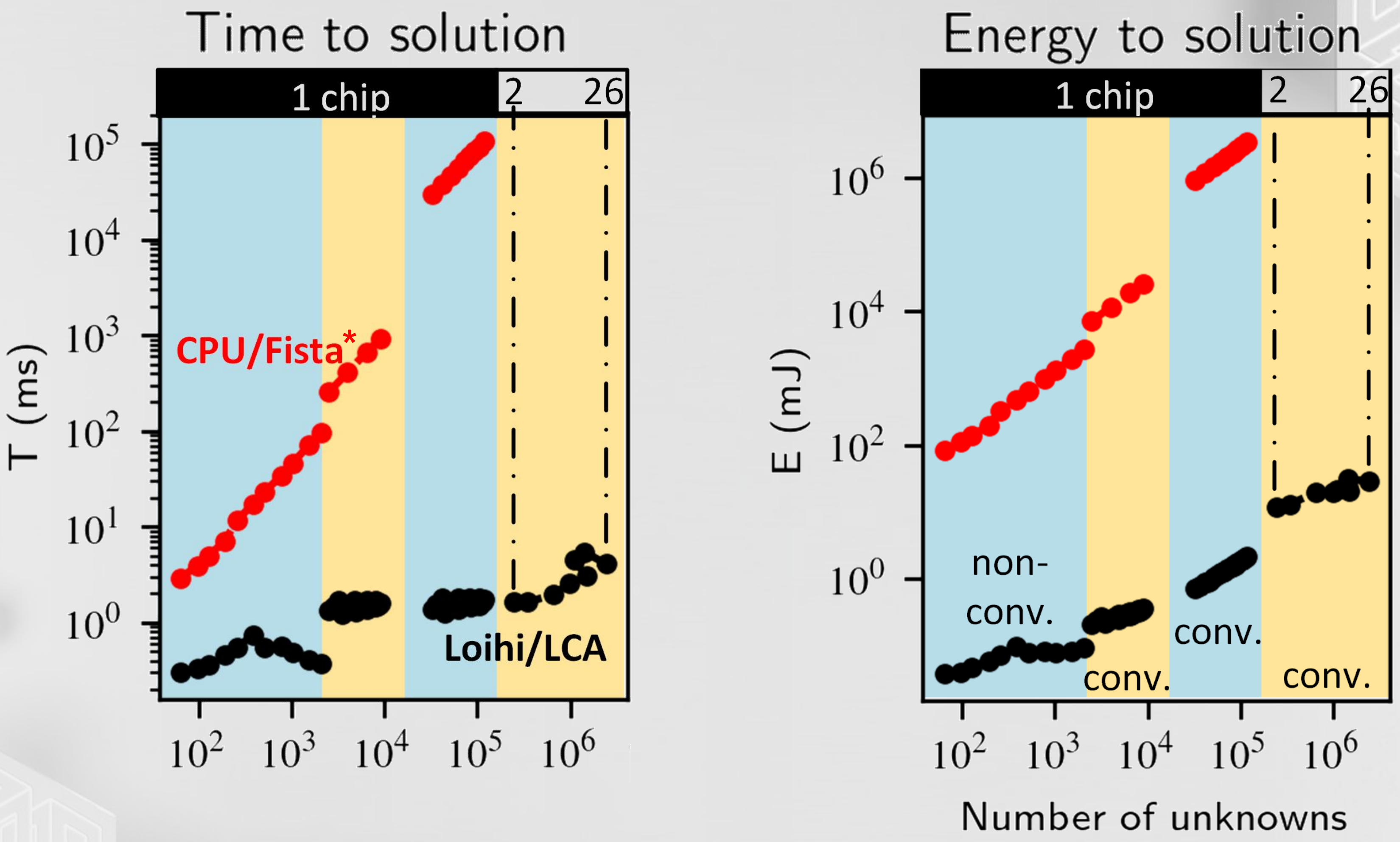
In the neural network formulation, **feature neurons compete** to reconstruct image with as few contributors as possible

Tang et al, arxiv: 1705:05475

Neural Network Structure

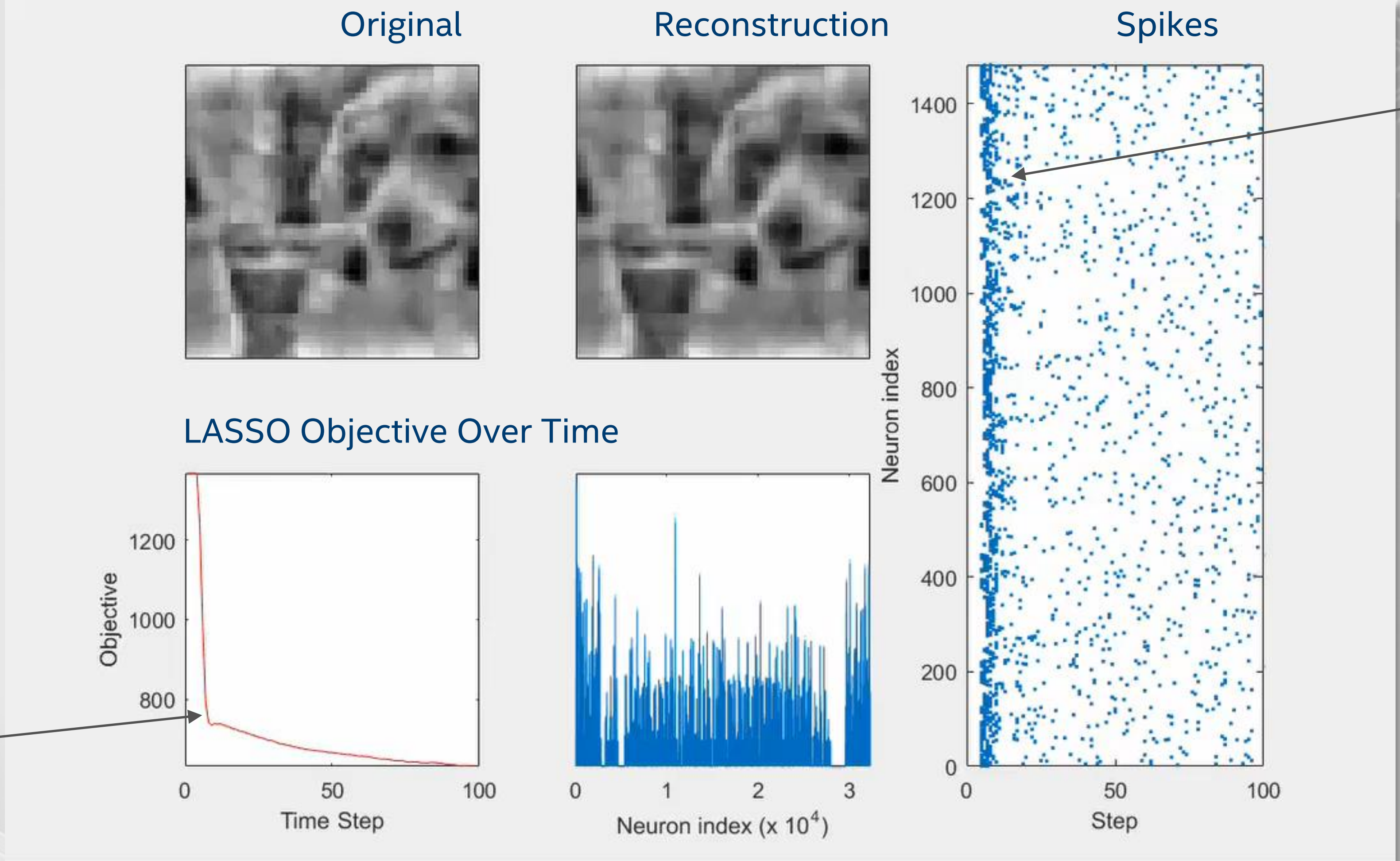


LCA Solver for LASSO Scales Incredibly Well on Loihi



* **Intel Core i7-4790 3.6GHz w/ 32GB RAM.** FISTA solver: **SPAMS** http://spams-devel.gforge.inria.fr/
Performance results are based on testing as of August 2019 and may not reflect all publicly available security updates. No product can be absolutely secure.

Spiking LCA Dynamics on Loihi



Intense but very brief period of competition

Rapid convergence on a neuromorphic architecture

Great efficiency comes from exploiting sparsity in space and time

Path Planning with Spikes

Runtime comparison to best Dijkstra optimizations:

- Neuromorphic: O($L \cdot \sqrt{V}$)
- Standard: O(E)

For most nontrivial problems:

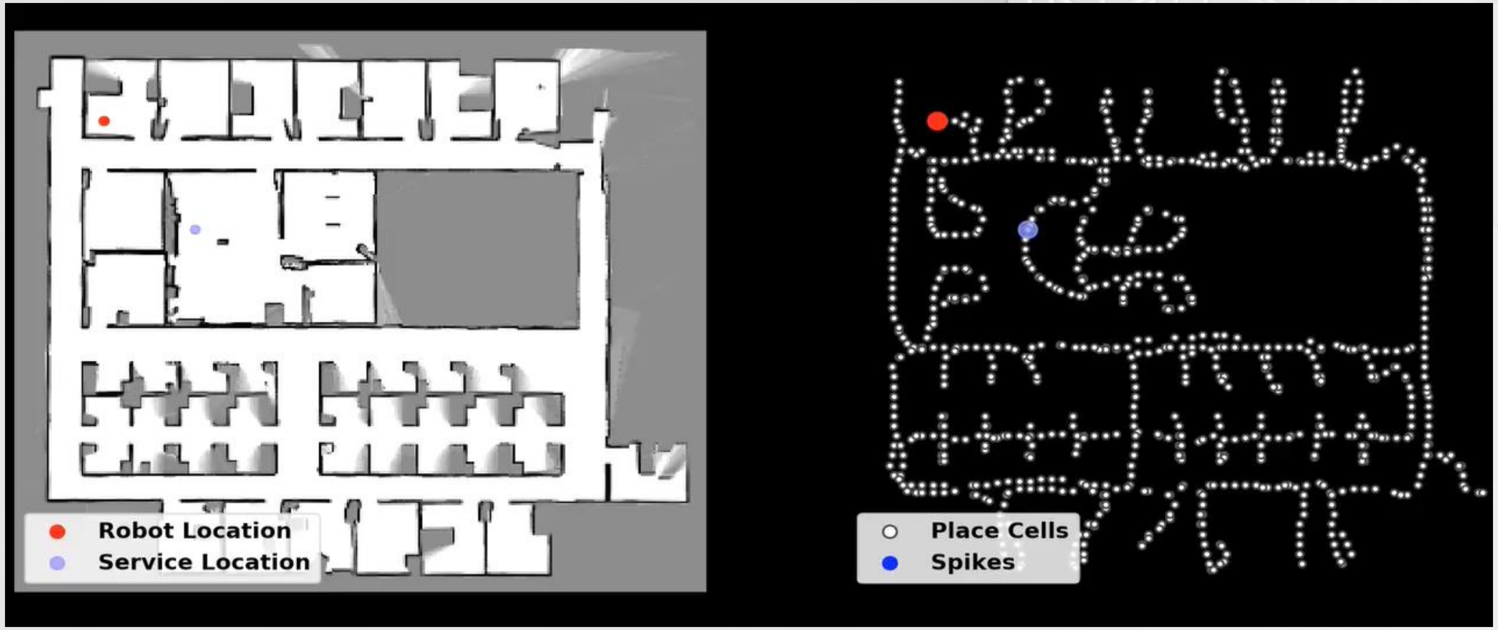
- L<<E
- V<<E

Neuromorphic solution uses *fine-grain parallelism* an *temporal wavefront-driven computation* to potentially provide great performance gains for large problems.

Based on *Ponulak F., Hopfield J.J. Rapid, parallel path planning by propagating wavefronts of spiking neural activity. Front. Comput. Neurosci. 2013. V. 7. Article № e98.*

ROBOT MOTION

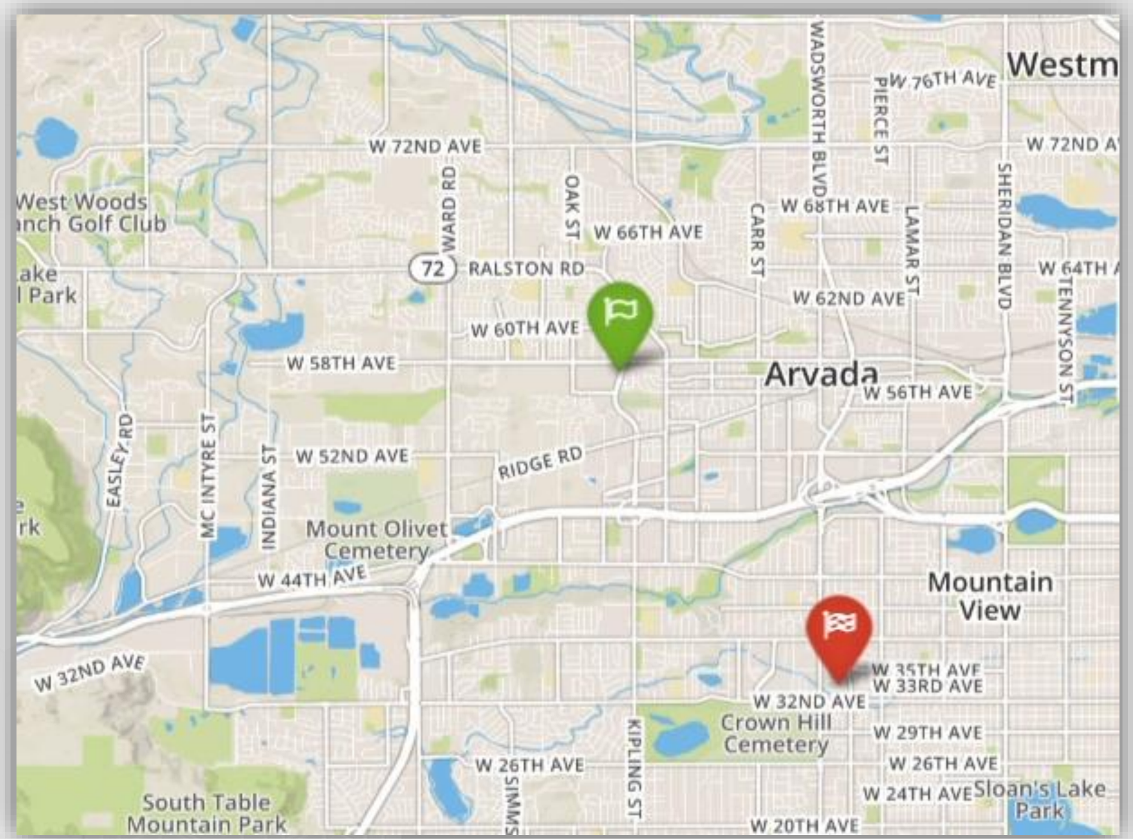
LOIHI REPRESENTATION



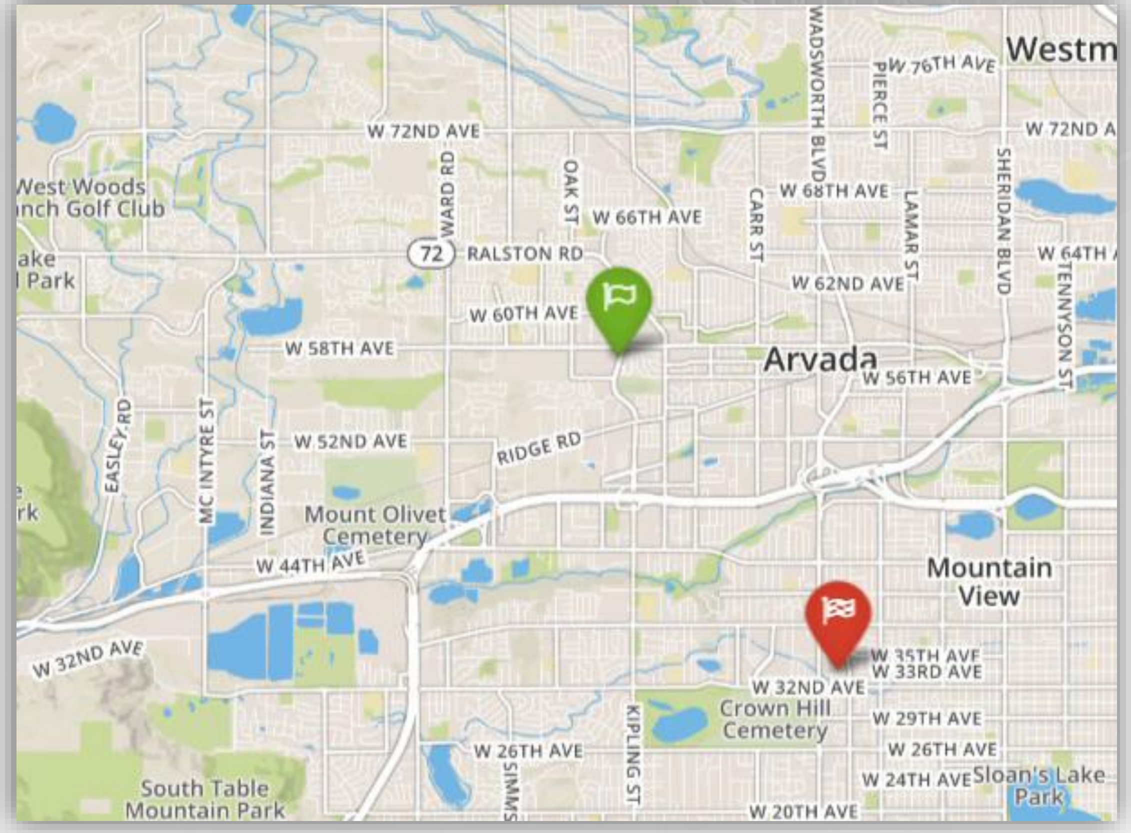
DARPA SDR Site B
(Data from Radish Robotics Dataset)

Using Loihi for Driving Directions in Colorado

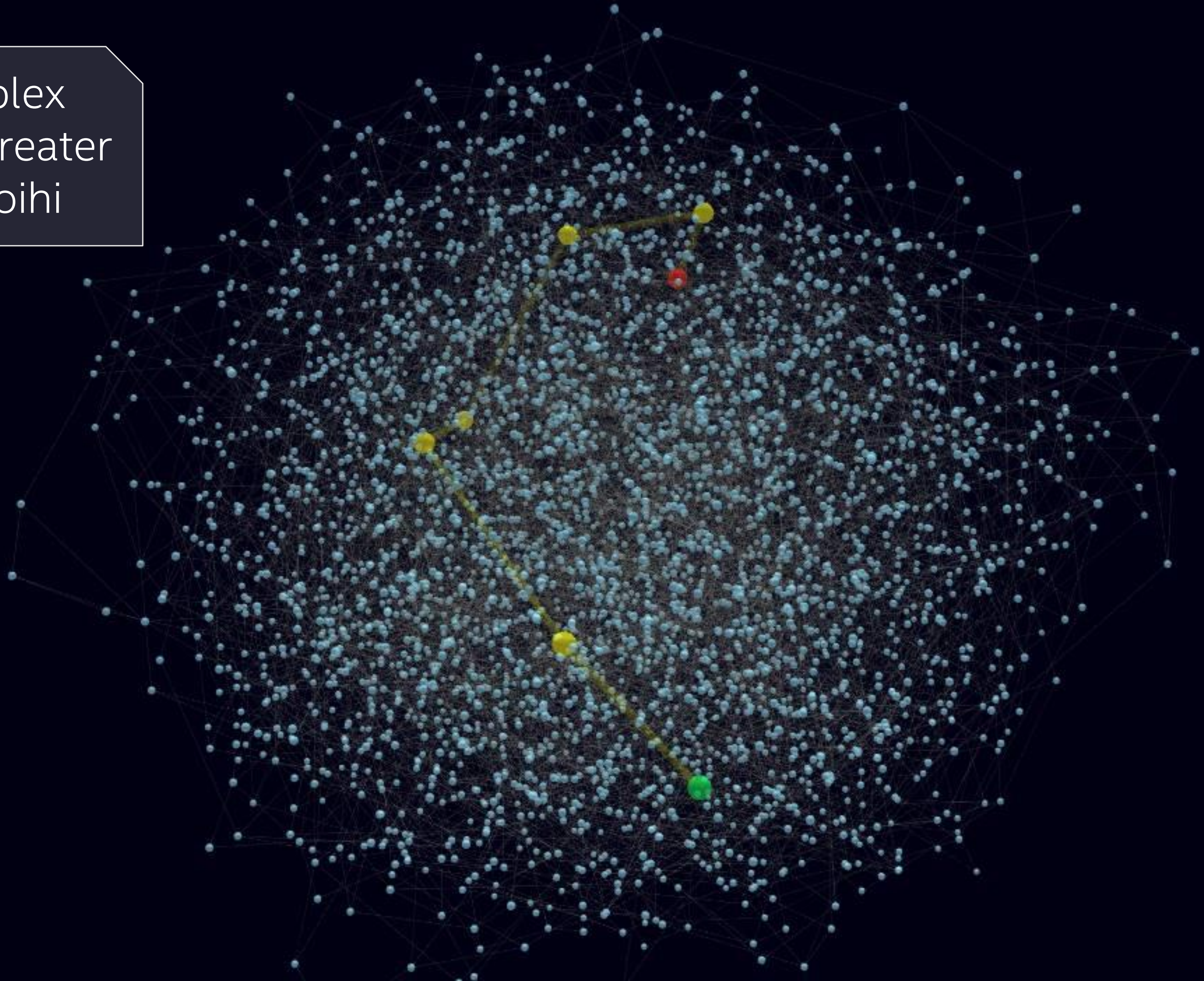
Loihi: Fine-Grain Parallel Search



Dijkstra: Sequential Breadth-First Search

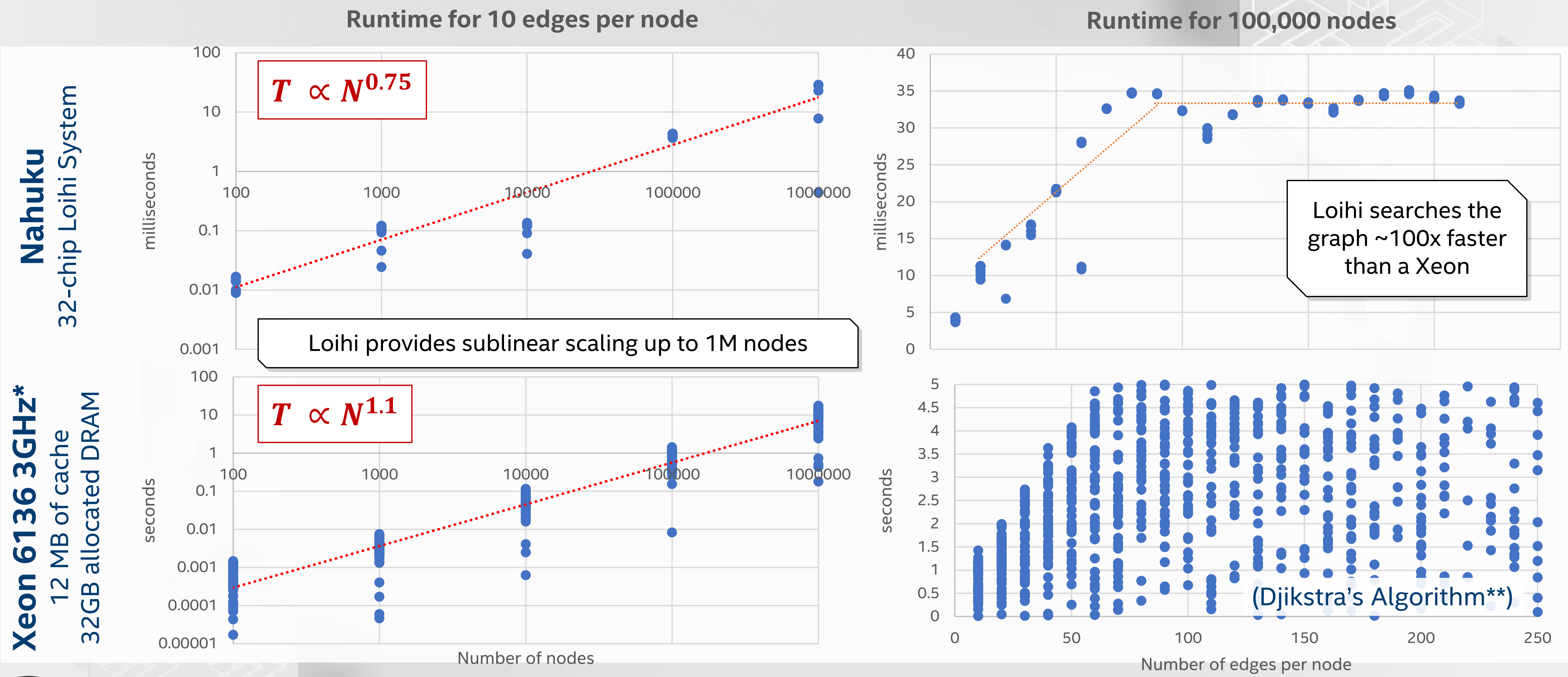


More complex graphs give greater gains for Loihi



Searching Small World Networks with Loihi

Watts-Strogatz network model with rewiring probability 20%.



* Intel Xeon 6136 3.00 GHz w/ 32GB RAM.

** with NetworkX graph analytics library

Performance results are based on testing as of December 2018 and may not reflect all publicly available security updates. No product can be absolutely secure.

The Research Frontier

Loihi is the **first neuromorphic chip** to demonstrate **compelling scaling results**
BUT THIS IS ONLY THE BEGINNING

FUNDAMENTAL PROPERTIES

- Fine-grain parallelism
- Local state
- Sparse temporal computation

ADVANTAGES

- Low energy
- Low latency
- Excellent scalability

Low Energy | Low Latency | Adaptive | Batch Size = 1 | High Cost

intel

Come See Pohoiki Beach **Live** in the Intel Booth!

Email inrc_interest @ intel.com for more information

Keyword recognition