

Intel Labs Announcement

INTEL NEUROMORPHIC RESEARCH COMMUNITY REQUEST FOR PROPOSALS (RFP)

SUBJECT

Intel's Neuromorphic Computing Lab (NCL) seeks proposals for neuromorphic computing research spanning theory, algorithms, applications, programming models, and sensors and controls. The research is expected to leverage Intel's Loihi neuromorphic architecture, using Intel's NxSDK software development kit and prototype systems. Research findings will be incorporated into future Loihi silicon iterations and may help lead to the commercialization of this technology.

Funding is available for eligible and suitably compelling proposals up to two years in duration. INRC corporate member **Accenture** also plans to support applications-oriented academic neuromorphic research and will independently consider proposals submitted to this RFP for funding.

KEY DATES

Upcoming INRC Talks and Information Sessions:

March 18, 2020. **INRC Forum**. 9:00-9:30am PDT.

April 15, 2020. INRC Forum. 9:00-9:30am PDT.

RFP Overview and Q&A sessions.

Please email <i>inrc_interest@intel.com to receive an online meeting invitation if you are not already an engaged INRC member.

For more upcoming events, see the INRC website:

http://neuromorphic.intel.com/

Email <u>inrc_interest@intel.com</u> for more information, including access to the INRC website.

Proposal Submission Deadline (PIs):

For proposals requesting funding beginning July, 2020: *April 8, 2020* For proposals requesting funding beginning Sept/Oct 2020: May 20, 2020 For all other proposals: *We will consider access-only project proposals at any time.*

OVERVIEW

In March of 2018, Intel's Neuromorphic Computing Lab launched the *Intel Neuromorphic Research Community* (INRC). This collaborative research program is open to all academic, government, and industry research groups interested in neuromorphic architectures for mainstream computing applications. The INRC will use Intel's neuromorphic research chip Loihi as the architectural foundation for development. Intel hopes the findings of this community will drive improvement of neuromorphic architectures, software, and systems, to accelerate the commercialization of this nascent technology.

For the majority of INRC activities, Intel expects its primary role will be to provide remote login access to Loihi systems and their accompanying software development tools across a range of engagement models. Intel is prepared to loan hardware and license development software to teams that require physical access for their proposed research.

Intel hosts regular workshops and meetings open to INRC project participants to share results, discuss challenges, and provide hands-on training. In addition to various minor events co-located with other conferences, Intel has hosted three major week-long workshops to date: in Reykjavik in Fall of 2018, in Portland Oregon in Winter of 2019, and in Graz Austria in Fall of 2019. The last was attended by over 100 researchers.

PROGRAM FUNDING

In the first round of INRC funding, which began in October of 2018, INRC academic members have received over \$1M in support from Intel's Corporate University Research Office (CUR). Intel is now inviting proposals for 1- to 2-year projects that address one or more of the research vectors (RV) RV1 through RV5 outlined in this RFP. Based on available budget, Intel expects typical grants to support one student or postdoc per project. All proposals must justify the proposed budget in terms of the resources needed to carry out the proposed work.

Additionally, Intel will share submitted proposals with **Accenture**. Accenture has allocated funds in support of applications-oriented academic neuromorphic research (primarily RV3 and RV4) and will independently consider funding proposals submitted for this RFP. Any groups that do not wish for their proposals to be shared with Accenture may opt out by indicating so in their proposal.

Due to the limited grant funds available, we highly encourage researchers to seek funding for their INRC projects from other sources by leveraging INRC support. Proposals outlining specific intentions to pursue such leveraging opportunities will be considered favorably. Upon request, Intel can offer letters of support for grant proposals to third party agencies. Feel free to submit such proposals to us at any time, as needed to support third party submission deadlines.

BACKGROUND

Computer architects and machine learning researchers have often looked to the brain for inspiration. Some of the earliest computer architectures and neural network models were developed with crude models of neural computation in mind. This includes the Von Neumann architecture that has thrived over many decades, as well as the Artificial Neural Network model that has more recently begun to deliver great practical value. At Intel we believe that now, more than ever, breakthroughs in computing and Artificial Intelligence may come from the study of neuroscience. We seek to apply the principles of neural computation as understood today to the form and function of silicon integrated circuits. In so doing, we hope that these *neuromorphic architectures* may better realize the brain's energy efficiency compared to conventional architectures and may match the brain's ability to learn and adapt.

Neuromorphic architectures are suited for a novel class of algorithms that we expect to be essential for ongoing progress in Artificial Intelligence. These algorithms promise to support behaviors more commonly associated with the brains of organisms than the programmed execution of computers. They will thrive in an environment of real-world stimuli, responding to environmental change with quick adaptation and rapid decision making. They will draw probabilistic inferences based on the accumulation of evidence. They will adapt on a variety of timescales, from quick one-shot events to more gradual changes accruing in response to unstructured data acquired over time.

Over the long term, we see promise for future neuromorphic processors across the entire spectrum of computing devices and form factors. At small scales, neuromorphic processors offer to give edge devices and sensors intelligent and adaptive information processing capabilities that today are far out of reach. For example, imaging sensors and surveillance cameras may perform preliminary saliency analysis and feature detection operations in the sensor itself before waking the more power-intensive vision subsystem.

We expect neuromorphic processors to be especially well suited for robotics and autonomous vehicles where real-time processing of multi-modal sensor data is required, followed by high-order planning, quick decision making, and low-latency output actuator control. All of these behaviors are well supported by biological brains and some of the specific relevant neural mechanisms in the brain are well understood and amenable to neuromorphic application.

At the high end of computing, the extreme scalability of biological neural architectures gives us confidence that neuromorphic architectures will similarly scale up to High Performance Computing (HPC) systems adept at solving massively complex high-dimensional analytics problems. *In-situ* data analytics is one compelling example application, where large volumes of data generated by HPC simulations must be analyzed in real time and specific patterns or anomalies need to be recognized for visualization purposes or for steering the simulation.

On the other hand, we do not necessarily see neuromorphic architectures as being the optimally efficient architecture for implementing conventional Deep Neural Networks (DNNs). DNNs solve a particular class of problems very well, and today's production chips such as Intel's Xeon® and Core™ processors, Movidius™ accelerators, and of course GPUs provide excellent solutions for those neural network models when sufficient quantities of labelled data are available. Neuromorphic architectures—spiking neural networks (SNNs) in particular—implement a broader range of

computation than conventional artificial neural networks, and we see many compelling applications of SNNs falling outside the domain of deep learning.

TECHNICAL OBJECTIVES OF RESEARCH

Since neuromorphic computing entails nothing less than a bottom-up rethinking of computer architecture, beginning potentially at the device technology level, unsolved and important research problems can be found at all levels of the computing stack, from the device technology level to circuits to microarchitecture and silicon design to algorithms to software programming models to end applications. Even the fundamental computing abstractions are open to debate, and some continue to question the value of spike-based models for practical applications. Figure 1 illustrates the full range of neuromorphic computing research vectors Intel and others are engaged in. Red vectors indicate hardware-oriented areas; blue vectors indicate theory and usage-oriented areas.



FIGURE 1. NEUROMORPHIC RESEARCH VECTORS

At Intel, we have developed Loihi, a spiking neural network research chip with an unprecedented range of features developed in a standard 14nm CMOS process, in order to thoroughly explore and validate what we see as the most promising neuromorphic architectural vector. For now, we choose not to consider the complexities of new devices or analog circuits. It is our view that, with the availability of Loihi, the state of neuromorphic hardware capabilities now significantly leads the state of algorithmic, application, and programming understanding. Therefore to efficiently accelerate progress in this field, we wish to defer explorations of alternate device technologies, circuits, and architectures.

Hence the focus of INRC and this RFP is on research vectors RV1 through RV5. For the foreseeable future, we intend to use Loihi and future silicon iterations as our primary vehicle for collaboration with a broadening network of researchers. As such, as a condition for Intel support, we ask that all project and grant proposals relate their objectives to the Loihi architecture and, more generally, the SNN computing model. More specific perspective and guidance for each research vector are provided in the sections that follow.

Given the state of neuromorphic research and our reading of priorities, we expect INRC activity and CUR funding investments to focus on RV2 and RV3, although we welcome projects in all research vectors.

RV1: Theory

As a new computational paradigm that differs from conventional computer architectures in fundamental ways, neuromorphic computing today lacks adequate theory in many important areas. A few of these are described below.

- Computational complexity frameworks. Conventional computational complexity analysis focuses on quantifying the number of primitive computational operations required to complete a given task, with data and instruction communication costs implicit. In neuromorphic architectures, communication and temporal staging of operations matter far more for performance and energy than raw operation counts. Relationships between sparseness of data encodings and performance/energy tradeoffs are fundamental yet poorly understood today.
- Characterization of neural dynamics. Neural activity is fundamentally dynamic and a wide range of dynamical behaviors are observed in the brain, e.g. coupled oscillations, fixed point firing rate attraction, stochastic attractors, chaotic and heteroclinic trajectories. Much work remains in order to abstract these behaviors to practical computational frameworks that may help classify and characterize the dynamics of spiking neural network algorithms.
- Quantification of neural engineering tradeoffs. The structure and function of biological brains have evolved to satisfy basic objectives of nature, such as minimizing energy consumption, response latencies, and production costs. By abstracting these properties and framing them in the context of invariant system objectives, we can more readily transfer the emergent design insights from natural brains to neuromorphic systems.

RV1 spans a very broad and unconstrained space of research. For INRC purposes, we will support projects pursuing this vector insofar as they align with practical application to the Loihi architecture or a space of clearly defined, relevant algorithms. We strongly encourage these projects to collaboratively apply their theoretical frameworks to other INRC projects.

RV2: Spiking Neural Network Algorithms

The meager availability of well-defined and functional SNN algorithms is the biggest impediment to progress in neuromorphic computing today. This especially applies to the domain of learning algorithms. We designed the Loihi chip to provide a sufficiently broad range of features to support the latest models researchers at the forefront of computational neuroscience are investigating with the hope of enabling rapid algorithmic exploration and breakthrough discoveries.

Within the broad umbrella of SNN algorithms research, we maintain an open mind to any and all compelling directions. We generally view the highest value directions being those that integrate real-time learning and adaptation, thanks to an SNN's inherent tendency to integrate and respond to data in a temporal, real-time fashion. That said, we hope to eventually demarcate a broad range of computations that have highly efficient SNN implementations, and not all of these will include the long-term adaptations commonly associated with learning.

Some example algorithmic categories where we see promise are provided below.

- Unsupervised and sparsely supervised learning
- Online learning without catastrophic forgetting
- Learning probabilistic models
- Learning through structural plasticity

- Reinforcement learning
- Sparse coding, especially with dictionaries that are learned online
- Bayesian inference, e.g. with Markov Chain Monte Carlo
- Associative memories
- Predictive coding
- Anomaly detection
- SLAM and autonomous planning
- Adaptive dynamic control algorithms
- Audio signal processing, e.g. as applied for speech and speaker ID recognition
- Object detection from event-driven image sensor technologies (e.g. Dynamic Vision Sensors)
- Optical flow computation

We encourage projects to explore emerging models of spiking/event-based computation that deviate from conventional rate-based formulations. The following themes are under study at Intel and are of particular interest:

- Computation with spike times
- Phasor-based computation with spikes
- Spike representation of data structures
- Spiking implementations of vector-symbolic computations

Equally important to us as the particular class of algorithm is the methodology used to develop it. We strongly prefer a principled, objective-driven mathematical approach to algorithm development. Without a mathematical model describing the dynamics of an SNN, we believe it becomes very difficult to abstract the network's behavior and parameterize its formulation such that it can be applied for generic computational use.

Also critical is the objective of supporting quantitative benchmarking of each algorithm's practical value. Algorithmic research projects should include the development of methods for quantifying the algorithm's value. These will commonly include resource consumption (e.g. Loihi neuron resources), algorithmic performance (e.g. classification accuracy), energy to solution, and time to solution. Whenever possible, these metrics should be compared to the corresponding values of the most analogous conventional algorithms evaluated on conventional architectures.

For INRC purposes, we are standardizing on the "algorithm" term to emphasize the goal of developing end-to-end computational procedures that process input and produce output in specified formats while performing some well-defined computation. Given a well characterized SNN algorithm, it should be a straightforward task to map this to a software implementation suitable for application use. RV2 projects should advance as far as possible beyond theory and specific experimental studies to yield generic SNN software modules that can be applied to a range of different problem domains.

RV3: Neuromorphic System Applications

For purposes of INRC and this RFP, system applications refer to specific software instantiations of interacting SNN algorithms demonstrated on Loihi silicon processing real-world data. These may be live demonstrations of Loihi silicon deployed in a physical system built by an INRC participant, such as an autonomous drone or robot, or they may be computational in nature, running on Intel's hosted neuromorphic servers while processing a static dataset. The overriding goal of this research vector is to demonstrate capabilities that are not possible to otherwise demonstrate with conventional solutions, whether due to novel functionality or power/performance advantages.

Benchmarking or, more generally, quantifying the value of each application demonstration is vital. No RV3 project proposal should neglect this requirement. In some cases, quantifying value may be as simple as measuring classification accuracy on a standard machine learning dataset. Where new functional capabilities are demonstrated, value quantification may require some dedicated research attention, depending on the nature of the capability.

Examples:

- Real-time adaptive control of autonomous robotic systems
- Object detection, classification, segmentation, and parameter estimation, e.g. in video streams and event-based data
- Signal extraction from noisy/unstable/fast-changing backgrounds (e.g. audio, video, telescopes, LIDAR, olfaction)
- Standardized interface to robotic systems and simulators
- Personalized speech recognition, translation, or speaker identification
- Vision-based gesture recognition, including learning new gestures online
- Tactile pattern discrimination with artificial skin technologies

Intel is prepared to provide physical access to Loihi systems and discrete chips in support of sufficiently compelling and credible RV3 project proposals. Such projects should plan for an initial phase of software development remotely using Intel's neuromorphic cloud service, with subsequent access to physical Loihi systems dependent on satisfying defined milestones.

RV4: Neuromorphic Programming Models

In almost all cases, programming neuromorphic algorithms today involves a very low level of structural programming. This requires specifying a spiking neural network in detail, including a precise graph connectivity and a long list of neural parameters. Support code may be developed in an ad hoc manner for data format conversion purposes and network management during the computation.

While an expert-oriented structural programming model may be adequate for today's immediate research needs and some specialized applications, we hope that higher level abstractions will emerge over time to improve the accessibility and productivity of neuromorphic hardware. Such abstractions might allow application developers to construct SNNs using symbolic, algebraic, behavioral, developmental, trained, or evolutionary representations. The new programming languages,

frameworks, and paradigms supporting these abstractions are the focus of this research vector. Over time we see breakthroughs in this area being essential for the mainstream adoption of neuromorphic computing systems.

RV5: Event-Driven Input/Output Interfaces for Neuromorphic Systems

This is the one vector of the RFP that may include a hardware component. Over the past few years, we have observed the maturing of event-driven imaging sensors that trace their roots back to Carver Mead's early work in neuromorphic engineering in the 1980's. These event-driven sensors inspired by the operation of the retina provide benefits in response to latency, power, and dynamic range over conventional imaging sensor technology.

Moreover, such event-driven (or spike-based) sensing strategies offer the most natural input data format for neuromorphic spiking chips such as Loihi. We believe many sensing modalities may benefit from such spike-based encoding strategies implemented at the lowest hardware level.

Event-driven actuator, display, and other system output functions have yet to receive much research attention. We see promise for similar latency and power savings in this domain and encourage research in this area. This includes interfaces to motors that support PFM, PWM, 3-phase based control, direct drive and servo motors, or muscle-like actuators, in conjunction with new event-based control algorithms.

Beyond funding, Intel will support this research vector by offering hardware support for AER-style input and output interfaces in our Loihi systems, and is open to supporting other protocols and interfaces that may arise from RV5 projects in future iterations of Intel's neuromorphic hardware.

OVERALL RESEARCH GOALS

The ultimate goals for INRC and Intel's neuromorphic research are the following:

- 1) Identify, develop, and characterize a space of algorithms that exploit the novel properties of spiking neural network hardware architectures (e.g. fine-grain parallelism, sparse activity, high degrees of connectivity) to deliver orders of magnitude gains in efficiency compared to the leading conventional algorithms.
- 2) Guide the iterative development of neuromorphic, non-Von Neumann architectures with the algorithmic and architectural insights that arise in the pursuit of goal (1).
- 3) Prototype real-world applications of Intel's neuromorphic silicon to assess the practical value that this architecture may provide if commercialized.
- 4) Develop an ecosystem of researchers and developers who can routinely and successfully apply Intel neuromorphic silicon to solve new problems.
- 5) Move neuromorphic computing from the fringes of AI and machine learning into the mainstream.

Project proposals that maximally align to these goals and articulate a strong and credible case for accelerating progress to these ends will be preferred.

ENGAGEMENT PROCESS

The general process for engaging in INRC research is outlined below.

- 1) Complete and submit one or more INRC project proposals
- 2) An *INRC participation agreement* needs to be executed between Intel and the institution affiliated with each member of a selected project's team. An authorized legal representative from the institution needs to sign this agreement, typically not the PI. Specific terms of the agreement may be negotiated in some cases, when necessary. For projects funded through Accenture, an agreement must to be executed between Accenture and the institution affiliated with each member of a selected project's team.
- 3) Once a project has the green light to proceed, members of the project formally become INRC participants. This allows you to sign up for a number of services:
 - a. INRC Participants site access, including detailed Loihi/SDK documentation and content contributed by other INRC members.
 - b. Remote access to Loihi via our Neuromorphic Research Cloud (NRC) system.
 - c. Access to the Loihi SDK, including software modules contributed by other INRC members.
 - d. Physical access to loaned Loihi hardware systems (as needed/approved)
- 4) Conduct your research project.
- 5) Share progress, results, and demonstrations in periodic INRC online forums and face-to-face workshops open to all INRC members.
- 6) At the conclusion of the project, you present a final report/demo, publish results, and contribute any shared software and/or code to the INRC software repository.
- 7) Having successfully completed at least one INRC project, you remain an INRC member but, due to capacity constraints, you may lose access to Loihi/NRC resources unless approved for another project.

Additional details describing the engagement process are provided in the sections that follow.

Eligibility

We welcome groups of all types and locations to submit research proposals and engage in the community, subject only to U.S. export control laws.

Intel research grants are only offered to academic research groups.

Project Proposal Submission

Along with this RFP, we are providing an *INRC Project Proposal Template* document that lists all information and documentation required of each respondent. Please refer to that document for detailed guidance on what information to include in your INRC project proposal, while preserving the template structure. A proposal may be rejected if it does not include the required information and documents.

Some recommendations:

- Delete all commentary and guidance text from the template document.
- Strive for brevity.
- Feel free to submit more than one proposal.
- Keep the scope of each project narrowly defined and limited to a single research vector.

Please note that Intel is unable to receive proposals that are under an obligation of confidentiality. All proposals submitted should therefore include only public information.

Progress Updates and Results

Once your INRC research project is approved and underway, we would like to receive quarterly updates on your progress and also notification of any work published as a result of this research. If you receive Intel funding, we will expect progress updates at least once per quarter, with at least one per year being an INRC face-to-face workshop.

Progress updates will generally be in the form of short \sim 30 minute presentations, open to all other INRC participants. Live demonstrations of results using Loihi, emulation, or simulation when possible will be highly appreciated.

Intel supports and encourages publishing results in public, peer-reviewed forums. We will do our best to support live demonstrations and independently hosted hands-on workshops using Loihi hardware systems, subject to U.S. export control, security, availability, and other constraints.

Loihi Hardware and SDK

Intel offers a wide variety of Loihi-based neuromorphic systems and software for you to utilize in your research.

Systems





Neuromorphic Research Cloud (NRC)

For some projects, the NRC may not work for you. Instead, you may need local access to our Nahuku-32 platform. While we have a limited supply, if you provide clear reasoning within your proposal for one, we can loan these systems for up to 3-6 months at a time. We have a smaller variant as well (Nahuku 08) which can be loaned for longer periods.

Our primary means of providing you access to Loihi is through our Neuromorphic Research Cloud. Approved projects will be given their own virtual machine on the





Kapoho Bay (KB)

Another system available for loan is our Kapoho Bay USB form factor. It supports up to 2 Loihi Chips and is compatible with Ubuntu 16.04 and 18.04. Great for portable projects and equipped with GPIO pins for DVS camera input, this device can be loaned to project teams for up to 1 year.

Software

We provide a complete software development kit to project teams. The kit provides a number of intuitive APIs, all Python native, allowing you to focus on your algorithms while we focus on getting them running on Loihi. A complete list of API options is provided below.

NxCore	This API provides basic Python register-level access to the Loihi hardware. Lacking high-level programming abstractions, the NxCore API is well suited for coding custom graph partitioning schemes but it may be too low-level for more complex application development.
NxNet	NxNet is our canonical API written natively in Python. This API allows you to think in terms of SNN abstractions such as groups, layers, neurons, and connections. It provides intuitive, programmable objects and employs the declarative (graph structured) programming paradigm.
NxTF	For mapping DNNs to SNNs we provide the NxTF API. This API looks and feels like Keras and enables programming and compiling typical DNN topologies for Loihi. Additionally, we provide a Loihi backend for the popular "SNN-Toolbox" framework to convert pre- trained ANNs into rate-coded SNNs.
NxSim and "improved API"	Coming later in 2020, we will be providing a functional simulator, coinciding with the release of our "improved API."
Nengo	Nengo, from Applied Brain Research, is an excellent tool for getting started with Loihi. Learn more at Nengo.ai.
INRC Contributed Frameworks	There is a growing list of contributed frameworks which are also available for use in INRC projects.

INTELLECTUAL PROPERTY

This solicitation affords proposers the option of submitting proposals for the award of a grant, a sponsored research agreement, or other agreement as appropriate. Intel and Accenture reserve the right to negotiate the final choice of agreement.

The final award terms are expected to follow one or the other of two high-level intellectual property (IP) approaches. Either: (1) Intel or Accenture and the university will jointly agree that IP developed under a grant will be placed in the public domain, including offering software under an open source license, or (2) Intel or Accenture and the university will negotiate a sponsored research agreement with more specific IP terms, which, at a minimum, will require the university to grant Intel and other sponsors (if any) a broad non-exclusive royalty free license to foreground IP.

It is a requirement to follow approach (1) if a project's software development directly enhances Intel's NxSDK code or builds on pre-existing INRC shared software libraries.

Please note that Intel and Accenture are unable to receive proposals under an obligation of confidentiality. All proposals submitted should therefore include only public information. Accepted proposals may be published to the INRC member site for community reference (*i.e.* visible to all other members engaged in INRC research), specifically sections 1-7. Groups will have control over all such content on the INRC website and may request for their project details not to be shared at all in this manner with other members, if so desired.

POINT OF CONTACT FOR INQUIRIES AND SUBMISSIONS

Proposal submissions and related inquiries should be sent to <u>INRC Project Proposals@intel.com</u>.

This RFP is administered by Intel Labs with Accenture's independent sponsorship of particular project proposals of its choosing. Interested groups are encouraged to contact <u>neuromorphic inquiries@accenture.com</u> for any questions on Accenture's selection process and involvement in this RFP.

LEGAL DISCLAIMERS

The issuance of this RFP and the submission of a response by a respondent or the acceptance of such a response by Intel Corporation ("Intel") or Accenture LLP ("Accenture") does not obligate Intel or Accenture in any manner. The RFP is not an offer or a contract. Intel and Accenture are not obligated to contract for any of the products/services described in the RFP. Intel and Accenture reserve the right to:

- 1) amend, modify or withdraw this RFP;
- 2) revise any requirement of this RFP;
- 3) waive any requirements of this RFP that are not material;
- 4) seek clarifications and revisions of responses to this RFP;
- 5) require supplemental statements or information from any responsible party;
- 6) accept or reject any or all responses to this RFP;
- 7) extend the deadline for submission of responses to this RFP or otherwise modify the schedule set forth in this RFP;
- 8) negotiate potential terms with any respondent to this RFP;
- 9) engage in discussions with any respondent to this RFP to correct and/or clarify responses;
- 10) require clarification at any time during the procurement process and/or require correction of responses for the purpose of assuring a full and complete understanding of a respondent's proposal and/or determine a respondent's compliance with the requirements of the solicitation; and
- 11) cancel, or reissue in whole or in part, this RFP, if Intel determines in its sole discretion that it is its best interest to do so.

Intel and Accenture may exercise the foregoing rights at any time without notice and without liability to any respondent or any other party for its expenses incurred in preparation of responses hereto or otherwise. All costs associated with responding to this RFP will be at the sole cost and expense of the respondent. Intel and Accenture make no representation or warranty and shall incur no liability under any law, statute, rules or regulations as to the accuracy, reliability or completeness of this RFP.