



INTEL REQUEST FOR PROPOSALS (RFP)

SUBJECT

Intel's University Research & Collaboration Office (URC) request proposals for academic research on "*Transformative Server Architectures*".

KEY DATES

Proposal Submission Deadline (PIs): July 6th, 2021

Response Date to Proposal (Intel): August 10th, 2021

Planned Projected Research Start: Targeting Q3/Q4 2021

OVERVIEW

Intel's University Research & Collaboration Office (URC) invites proposals to establish a new multi-university collaborative research center to develop technologies that enable Transformative Server Architectures (TSAs). TSAs for modern datacenters need to enable extremely cost and power efficient online services at low latency and high flexibility. The TSA center should deliver leap-ahead performance and enable a fundamental upward shift in the future trajectory of performance-per-watt and performance per Total Cost of Ownership (TCO) dollar for servers in scale-out datacenters.

Background and Problem Statement:

Single thread CPU performance gains for general purpose applications has slowed partly due to the slowdown of technology scaling. However, the demand for compute and very large volumes of data requiring compute continues to rise very rapidly. Meanwhile, application architecture in modern datacenters has evolved from monolithic to one based on microservices and FaaS (functions as a service). Such application decomposition has aligned with the server decomposition trends undertaken by Cloud Service Providers (CSPs). However, cloud-scale workloads in general (and microservices/FaaS in particular) introduce an "infrastructure tax" of wasted compute cycles due to increased CPU, memory, and network bottlenecks. Hyperscalers like Google, Facebook and Amazon are reporting 30-50% of cycles consumed by the infrastructure which must be drastically reduced to improve user experience and lower TCO. A focus on end-to-end system performance with a software and silicon architecture designed grounds-up and optimized for the modern datacenter is required.

To satisfy the at-scale requirements of modern massive online services, (CSPs) are building their hardware platforms using scalable fabrics with disaggregated components. These include general-purpose CPUs, heterogeneous accelerators (such as GPUs, ML accelerators and FPGAs), customized "infrastructure" or Data Processing Units (DPUs) and memory subsystems comprised of heterogeneous



memory technologies. This trend has come at the significant escalation of complexity and cost. The huge overheads of data movement and loss of overall compute and resource utilization due to cache, memory and IO bottlenecks has led to inconsistent CPU performance and system power inefficiency. The datacenter server architecture has evolved from where the CPU was the sole control point, to where the DPU has now become a credible alternative control point. Meanwhile, the Total Cost of Ownership incurred by CSPs has also steadily increased with the cost of Memory growing faster than that of Compute and other components.

To summarize, the overall datacenter server ecosystem is at an inflection point today. The proliferation of massive online services from the edge to the public clouds, on the one hand, and a disruptive change in technology with the introduction of silicon heterogeneity and scalable fabrics, on the other, call for a fundamental redesign of server stacks and systems.

Server Architecture Evolution:

Servers in the past two decades have gone through a major transformation from being the building blocks for High Performance Computing (HPC) in machine rooms to datacenters providing ubiquitous online services at a massive scale to serve individuals, enterprises, academic and research organizations, and governments. Today's datacenters rely on low-cost high-volume data processing, communication, and storage. Services are typically hosted in memory due to stringent microsecond tail latency requirement on service time and are often dominated by data management and in-memory analytics. Datacenter operators also offer cloud services through containers or Virtual Machines (VMs) to customers to maximize server utilization and return on investment.

The modern incarnation of a server blade employs a classical system organization with the CPU orchestrating computation over data in a memory hierarchy and the Operating System (OS) managing data movement from/to storage and the network. Scale-out datacenters are built with each blade server provisioned with sufficient memory to host a partition of the overall dataset. The aggregate provisioned memory accounts for a substantial and increasing fraction of a server's budget, and hence judicious efficient usage of memory directly translates to TCO savings. Moreover, movement across memory hierarchy levels is a key source of latency and energy not only for data management but also for data analytics. Finally, modern datacenter server software is layered with a rich and diverse functionality across layers. Efficient instruction supply to the CPU has thus become critical for high throughput performance and low tail latency in servers running deep software stacks.

Opportunities for Transformative Server Architectures:

The slowdown in silicon scaling, and the emergence of heterogeneous logic and memory technologies call for foundational server architecture research in several avenues to drastically reduce cost and power while scaling performance:

- There is a myriad of opportunities to innovate in CPUs. Innovation in both the instruction and data supply path is required to guarantee tail latency for services with deep software stacks. Fundamentally new CPU organizations for the Microservices/FaaS era should be explored.
- Advantageous integration of modern Non-Volatile Memory would require novel Terabyte-scale address translation mechanisms while maximizing performance and utilization of aggregate available memory (DRAM + NVRAM) and mitigating the impact of memory fragmentation.
- A careful multi-granular integration of domain-specific accelerators and FPGAs into the memory hierarchy and the CPU can efficiently scale compute throughput, data management, analytics and system services offloaded and controlled by a CPU host.



Grant/Gift External Coversheet University Business Office

- Network software stacks can also benefit from foundational CPU and DPU architecture, ISA, and microarchitecture innovation as communication software stacks (e.g. base networking, RPC, data transformation) becomes a bottleneck with scalable network fabrics in datacenters.
- Breakthroughs in fabrics that allow CPUs and heterogeneous processing engines to efficiently communicate and share memory with minimal data movement can enable superlative efficient performance for modern datacenter applications. Hence, we believe there exist huge opportunities for server TCO reduction from memory pooling and silicon consolidation.

To summarize, in recent years, with the emergence of heterogeneous accelerators, servers have witnessed a migration of computation from CPU-centric server nodes into discrete custom nodes. The strategic research objective of the TSA center is to place the CPU again at the center of a modern server holistically synergizing CPUs, heterogeneous accelerators, and memories through novel co-design down the stack spanning system software, architecture, and microarchitecture. The **overarching goal of the TSA research center** is to enable a 10x performance increase, and a 10X-50X performance-per-watt and performance-per-TCO-dollar leap over existing server architectures while offering significantly better programmability, flexibility, and scalability for future datacenters.

PROGRAM SCOPE AND FUNDING

Intel URC contemplates funding a new collaborative multi-university research center addressing all of the Research Vectors (RVs) below, described in detail in the next section.

- RV #1:** Datacenter CPU Core Architectures and Microarchitectures
- RV #2:** Terabyte-scale Memory Hierarchies
- RV #3:** CPU-centric Accelerator Ecosystem
- RV #4:** Applications and Workloads

This program intends to fund collaborative proposals, renewable annually at Intel's discretion. We request multiple PIs (can be from multiple universities) to team up to develop collaborative proposals for a research program that covers multiple RVs synergistically. It is recommended the collaborative proposal should include at least 2 PIs and collaboration across universities is strongly encouraged. Each collaborative proposal should aim for approximately \$400K of funding per year for 3 years.

One PI should be explicitly designated to lead the overall effort and coordinate among all the PIs involved in the collaborative proposal. Proposals should explicitly elaborate on how the PIs will collaborate and share research infrastructure across the RVs targeted. Intel reserves the right to form a coherent multi-PI academic research program that may span across multiple proposals in order to comprehensively address all RVs targeted by this RFP to meet its strategic objectives.

It is expected that a very high level of collaboration exists among PIs across all research vectors to develop experimental research prototypes that can work across RVs, and towards a unified capstone project to deliver full-stack research prototypes. The proposal should explicitly point out how each RV is addressed, the synergy among them, the plan and milestones towards building research prototypes, and the anticipated final full-stack prototype outcomes. The roles and budget for each PI should be clearly defined in the proposal, including justification and details of proposed budget in terms of the resources needed to carry out the proposed work, as well as milestones to achieve planned results.



Intel will be deeply engaged with the center and will assign partner technologists/collaborators across RVs to interact with the academic community to produce a stream of innovation proof-points, publications, demonstrations, and technology transfers into Intel and the broader industry throughout the duration of the program. We aim for the interaction to be bi-directional where Intel collaborators are part of the research team. Not only will they provide research feedback, but they will also actively contribute and co-develop the research to amplify the center outcome and enable continuous technology transfers into Intel and the broader industry.

TECHNICAL OBJECTIVES

Intel seeks to drive server CPU and platform architecture evolution while intensively engaging the overall datacenter ecosystem. Top academic groups around the world are researching avenues for server software-hardware co-design with novel abstractions for efficient server scalability. Intel seeks to leverage this rich faculty expertise and datacenter ecosystem connect through the TSA center. As noted earlier, the overall goal of the TSA center's research is to enable increases of 10x in performance and 10X-50X performance-per-watt and performance-per-TCO-dollar over current server architectures.

Architecture Research

Four vectors of research (RV1 through RV4) for the center are defined as follows:

RV #1: Datacenter CPU Core Architectures and Microarchitectures

Modern cloud applications heavily use languages such as Java, Python and Node JS. Such codes are often interpreted or JIT-compiled, dynamically typed, highly threaded and automatically garbage collected. They execute many more instructions, have high memory usage, and use smaller basic blocks with high branch density and exhibit diverse code and data access patterns. Servers today also suffer from high frequency context switches and high address translation overheads as part of highly virtualized container-based execution. Novel CPU processor organizations and microarchitecture designs that are uniquely built to excel in modern and future datacenter execution environments should be the focus in this RV. Specific challenges to address and directions to research could include:

- A key challenge in modern server software stacks is instruction supply to CPU due to multi-megabyte server instruction working sets and hard-to-predict code patterns. This requires both accurate predictors for lookahead and dramatically larger metadata for predictor and Branch Target Buffer (BTB) tables, motivating datacenter application-centric predictor and BTB design.
- Since the instruction supply and corresponding metadata is often shared between concurrently scheduled threads on a modern datacenter CPU, many-core architectures with shared front-end instruction supply that feeds clusters of back-end lower complexity cores to maximize throughput per watt could be explored. Techniques to achieve out-of-order performance at in-order complexity through hybrid pipelines would be high interest in this context. The right level of simultaneous (hardware) multithreading (SMT), optimal pipeline implementation and QoS policies would be critical to get the right balance between throughput, latency, and efficiency.
- Dynamic profiling techniques that leverage the latest Intel® Xeon® CPU performance counter mechanisms can be used to dynamically learn execution bottlenecks and identify hot control-flow paths. Dynamic code re-layout and novel instruction prefetching with co-designed cache hierarchies could unlock significant cloud/datacenter CPU and system efficiencies.
- Advanced hardware-optimizers for CPU instruction elimination through dynamic multi-instruction fusion and dead code elimination in hardware through speculation or other mechanisms have significant potential to improve the energy efficiency and latency of today's online services.



- Novel virtual memory translation architectures and cache hierarchies that drastically reduce memory address translation overheads and improve cache utilization in modern virtualized environments and demonstrate scalability to large core counts on emerging applications with very large memory footprints are of high interest.

RV #2: Terabyte-Scale Memory Hierarchies

A key challenge in today's datacenter servers is providing "just-in-time" memory capacity with an effective negotiated memory latency. Emerging memory technologies such as 3D die-stacked DRAM and non-volatile random-access memory (NVRAM) can enable high performance. Server designers are increasingly combining traditional and emerging memory technologies. But such heterogeneous memory hierarchies exhibit non-uniform access performance characteristics.

Intel's byte-addressable Optane™ Terabyte-scale NVRAM introduces significant opportunities such as extending memory capacity and persistence. It is well-suited to many high memory footprint applications such as in-memory databases and big data analytics. In App Direct Mode, DRAM and Intel® Optane™ DC Persistent Memory are both accessible as total platform memory, whereas in Memory Mode, DRAM is used as a hardware managed cache. However, Terabyte-scale address spaces stress translation mechanisms not only for accelerators but also for modern CPU's. Terabyte-scale memory hierarchies also present challenges such as increased hierarchy depth and longer latency and requires novel programming abstractions. Large memory footprint server applications using terabytes of memory incur frequent TLB misses.

Key objectives in RV #2 could include:

- Ultra-low overhead next-generation Terabyte-scale address translation architecture providing efficient TLB reach for both CPU address translation (MMU) and accelerator address translation (IOMMU), co-designed with system software and OS.
- End-to-end cache and memory compression to minimize memory fragmentation while aggressively optimizing tail-latency of emerging services. Translation and compression architecture should be well integrated with memory movement decisions across the hierarchy.
- a) Support for application-transparent data migration on heterogenous memory architectures that includes mechanisms to detect and track hot and cold data, spawning software-only, hardware-only or hardware-software co-designed approaches. b) Seamless and optimally architected multi-level heterogeneous memory hierarchies with application and system software abstractions for memory placement and data/metadata migration. Mechanisms as in a)/b) need to be intelligently co-designed with data placement and movement decisions across the hierarchy, including novel processor(hardware)-initiated multi-level prefetching managed across cache controllers in the CPU and the memory controllers in the system. Simultaneously benefiting from both from NVRAM's high capacity and DRAM's performance, while drastically lowering datacenter-level TCO for next generation services should be a central goal in this RV.
- Novel Near-Memory-Processing architectures to drastically improve datacenter server performance and power efficiency while employing heterogenous terabyte-scale memory hierarchies

All research objectives suggested above should comprehend the opportunities afforded by Terabyte-scale memories towards pooling/sharing and address the challenges thereof.

Researchers are also encouraged to study big memory algorithms that tradeoff memory capacity and compute. For instance, with Terabyte-scale memory capacity, could key algorithms be built differently? What is the ideal memory bandwidth/capacity ratio for emerging application domains?



RV #3: CPU-centric Accelerator Ecosystem

Servers in datacenters today map diverse applications onto the CPU and a heterogeneous mix of accelerators such as GPUs, domain-specific accelerators, FPGAs, and Data Processing Units for infrastructure tasks. The CPU and the OS are the entry points to host services, and they coordinate compute offloads to accelerators. However, a wide spectrum of accelerators in today's datacenters are separated from memory through legacy OS abstractions and IO fabrics. A CPU-centric accelerator ecosystem can be a fundamental enabler to effectively share memory, avoid silicon fragmentation and reduce customer TCO. Further, emerging network fabrics allow for aggregating rack-scale memory with microsecond turnaround for paging. But their advantageous adoption requires inter-server load/store semantics, address translation and protected data movement co-designed with network controllers.

Key objectives in RV #3 could include:

- User-level programming abstractions, memory models and coherence fabrics for a seamless shared-memory based CPU-centric accelerator ecosystem enabling tight orchestration of modern datacenter workloads on heterogeneous accelerators. This should span aspects such as efficient communication and signaling and low-cost virtual/unified shared memory and memory tiering supporting both host and device-attached memory.
- Near-Memory-Processing for heterogeneous computing with seamless integration of accelerators into the CPU's shared address space with isolation guarantees. New fine-grained and coarse-grained compute primitives for near-memory-processing that take advantage of the increased bandwidth and lower latency available closer to memory for data intensive workloads. Novel hardware abstractions to exploit near memory processing compute in the SOC, including address translation, interfacing, and signaling.
- Scale-out CPU fabrics that provide efficient message and load/store semantics to applications across compute nodes with strong isolation properties while bypassing communication libraries and network stacks, including new architecture or ISA support for addressing remote memory properties (e.g., availability/failures, memory ordering). New peer-to-peer (P2P) IO fabrics that enable high-performance P2P traffic while maintaining security at the target accelerator.

Since the end-to-end performance of application workloads in modern distributed, multi-user, and data-centric contexts depends heavily on system software efficiency, we invite researchers to explore systems-level acceleration schemes made possible by new TSA architecture innovations.

RV #4: Applications and Workloads

Microservices/FaaS applications incur significant overheads owing to their application architecture and distributed deployment. Such overheads include instruction cache misses, instruction TLB misses, CPU front-end stalls, and those related to virtual memory management, data serialization and deserialization, network stack processing, low-power states, and thread scheduling. Compute cycles not used for application processing lead directly to loss of revenue for CSPs. The performance overheads of Microservices/FaaS deployments are amplified in multi-instance environments when hundreds or thousands of concurrent functions corresponding to independent instances are competing for limited hardware resources. This drastically increases CPU stalls and restricts the scale-out that a CSP can provision on a physical system resulting in under-utilized datacenters and higher TCO.

Hence, full-stack characterization of state-of-the-art microservices/FaaS-based and other representative modern datacenter applications on the latest server platforms, and identification of key bottlenecks across application code, runtimes, system software (OS) and the hardware domains is an essential initial task for TSA research teams. Detailed analysis of CPU bottlenecks using Intel's top-down methodology cross the entire processing pipeline (spanning branch prediction, instruction fetch and



Grant/Gift External Coversheet University Business Office

decode, micro-op allocation and dispatch, address translation, interconnect congestion, and cache and memory/IO bottlenecks) can be particularly illuminating in setting specific next-level architecture research objectives.

Researchers are invited to consider emerging applications that exploit key TSA innovations in next generation server architectures for future datacenters. These can stretch from algorithmic kernels to end-to-end system-level applications representative of realistic deployments. TSA center researchers are strongly encouraged to explore futuristic datacenter cloud applications and scenarios for their server processor and system architecture innovations.

To achieve TSA's strategic research objectives, researchers are also strongly invited to a) develop benchmark suites that reflect real-world cloud-scale datacenter execution environments and b) identify a family of emerging key applications in target domains and their enabling algorithms. The suites may be comprised of reference algorithms and corresponding application software usable by the center (and the broader community) to better understand and quantify the gains of TSA innovations. The benchmark suites should extend to complete end-to-end applications that include realistic system and software stack components such as Operating System kernel, virtual machines, and containers. TSA researchers are expected to open source all such collateral developed.

Tools and Infrastructure:

Researchers are expected to evaluate applications and benchmark suites on Intel® Xeon® CPU datacenter servers with large capacity Intel® Optane™ memory, Intel® Xe or higher class of GPUs, and supporting high-speed networking and storage. Researchers should make use of common open source building blocks in demonstrating their approach with realistic system and data center software stacks (e.g., Linux, Kubernetes).

Specific methods and techniques of TSA research need to be evaluated using a combination of real systems, hardware emulation, and hardware/software simulation.

- Emulation platforms can be valuable for quick ramp into the research and idea evaluation.
- Picking the right simulation infrastructure at the level of abstraction that aids in a systematic deconstruction of the fundamental bottlenecks and inform and inspire innovations across RV1, RV2 and RV3 should form an essential part of the research planning. Substantial architecture-level modifications of server processor, cache and memory hierarchy through TSA research should be evaluated on RV4 workloads using appropriate (micro)architecture-level single and multi-socket system simulation using full software stack. The simulators should have high-fidelity, and span compute, fabrics, memory, and networking components of servers deployed in modern datacenters.
- RTL-level simulation and area/power/performance demonstration of critical modules from TSA research would be useful for effective technology transfers to Intel and the rest of the industry.

Intel expects its in-depth OS expertise and broad and deep engagement with the Linux community to significantly help proliferate TSA innovations that involve hardware-software co-design.



PROPOSAL FORMAT

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

Proposals should be up to 15 pages, not including citations or cost volume. (Proposals are encouraged to be succinct, but larger team proposals may exceed this page limit if absolutely necessary to describe all parts of the research.) Each response should comprise the following sections:

- **Cover page {1-2 pages}**. Title of proposal, name(s) of author(s), contact information, name of university, funds requested, the amount of cost share (if any)
- **Executive summary {1-2 pages}**. Define the technology that this research will develop, the performance targets that are expected or could be achieved, and the basic proposed approach.
- **Background and details of proposed technology {3 pages or fewer}**. This section is the centerpiece of the proposal. It should clearly and succinctly describe proposed technology in detail relative to the existing state-of-the-art in terms of performance, cost, and manufacturability. Describe the basis for the proposed technology performance targets, including any calculations, simulations, experimental data, or material properties that support those targets. Be sure to detail the current state-of-the-art for the proposed technology (or nearest related technologies). This section must also include an explicit statement of the Intellectual Property (IP) status for any and all background IP related to this technology (i.e., are the property rights to this technology protected, and if so, who owns those rights).
- **Detailed technical rationale, approach, and research plan {3 pages or fewer}**. Identify all aspects of the proposed technology that will require significant research or engineering development work in order to commercialize the technology. Describe in detail which aspects of the research and development needs for the proposed technology will be addressed by the proposed research plan (not all required research and engineering development work for the proposed need necessarily be addressed in the research plan, but they should at least be identified). Describe the scope, breadth, and purpose of the proposed research plan and how it will result in a viable commercial technology.
- **Statement of work, schedule, milestones, success criteria and deliverables {3 pages or fewer}**. For each of goals addressed, outline the 3-year scope of the effort including tasks to be performed, schedule, milestones, deliverables, and success criteria. It is understood that aspects of this research effort may be exploratory in nature and schedules/deliverables reflect intentions rather than a firm commitment.
- **Proposal team {2 pages or fewer}**. Summarize the members of the program team, their qualifications, and their level of participation in the project.
- **[Optional] Diversity and Inclusion (<1 page)**. In light of Intel's strong commitment to diversity and creating an inclusive environment, please address: (a) your organization's commitment to diversity and inclusion with respect to race, national origin, gender, veterans, individuals with diverse abilities and LGBTQ, and (b) a summary of your performance in this area and any initiatives you are pursuing.
- **Citations {unlimited}**.
- **Cost volume {unlimited}**. Cost proposal in Excel or other format as appropriate.



EVALUATION CRITERIA

In order of importance, the evaluation criteria for this solicitation are as follows:

1. **Potential contribution and relevance to Intel and the broader industry:** The proposed research should directly support a technology solution that addresses the RVs in the Technical Objectives outlined above, leading to technological advances with the potential for ongoing technology transfer in collaboration with Intel and the broader industry.
2. **Technical innovation:** Proposed solutions of interest should clearly push the boundaries of technical innovation and advancement. Research that is not of interest in this program include incremental advancements to state-of-the-art and current design practices.
3. **Clarity of overall objectives, intermediate milestones, and success criteria:** The proposed Research Plan should clearly convey that the PIs have the knowledge and capability to achieve the stated research goals. It is understood that any research program will have uncertainties and unanswered questions at the proposal stage, but a clear path forward in key challenge areas must be identified and justified. Teams are expected to demonstrate progress toward project goals at quarterly milestones and monthly project status updates. As detailed in “Program Scope and Funding” section, the proposal should explicitly point out how each RV is addressed, the synergy among them, the plan and milestones towards building research prototypes, plan for ongoing technology transfers, and the anticipated final full-stack prototype outcome. Strength of project management plan will also be considered.
4. **Qualification of participating researchers.** The extent to which expertise and prior experience bear on the problem at hand. Please elaborate on track records of building research prototypes and resulting publications from past relevant projects.
5. **Cost effectiveness and cost realism:** The extent to which the proposed work is both feasible and impactful within the proposed resource levels will be examined.
6. **Potential for co-funding.** Opportunity for closely synergistic matching grants and co-funding with other funding entities, such as SRC, NSF, DARPA, NSERC, EU etc. will be given significant consideration.
7. **Potential for broader impact.** Intel supports the advancement of computing education and diverse participation in STEM. Significant consideration will be given to proposals in which the outcome of the research can influence the development of new curriculum initiatives impacting undergraduate or graduate education at the respective universities (e.g. exposure to latest industry technologies/tools in classroom setting). Proposals are encouraged to elaborate on how the proposed work is anticipated to impact student education on campus and/or the broader academic community.

Intel Note: As an industry leader, Intel pushes the boundaries of technology to make amazing experiences possible for every person on earth. From powering the latest devices and the cloud you depend on to driving policy, diversity, sustainability, and education, we create value for our stockholders, customers, and society. Intel expects the suppliers in our supply chain to be strong partners in making Intel successful through support of Intel's goals and commitments to diversity, sustainability, and education.



PI MEETINGS AND COLLABORATION STRUCTURE

It is expected the PIs and student researchers will collaborate on a daily or weekly basis. Monthly PI, student and industry collaborator meetings will be used to review research results, present significant updates, and provide feedback and extend collaborations.

Semi-annual face-to-face or virtual meetings will be held to facilitate program-wide information exchange, review, and discussion of research. Researchers should anticipate one annual face-to-face meeting to be held at an Intel and one annual face-to-face meeting to be held at a university associated with this center. Associated travel costs for two annual meetings should be considered and included in the proposed budget. In the event unexpected travel restrictions prohibit a face-to-face meeting, a virtual meeting will be held.

To aid in collaboration across projects within the center and communication of research findings to the public, it is anticipated that a center website will be established, hosted, and maintained by the designated leading university. Intel requests the right to host the associated website link on their respective university program websites.

ELIGIBILITY

This RFP is open only to academic researchers and institutions that have been specifically invited to participate in the proposal process. However, invitees may freely select additional academic collaborators. Any questions regarding eligibility should be directed to Jeff Parkhurst (jeff.parkhurst@intel.com) and Sree Subramoney (sreenivas.subramoney@intel.com).

INTELLECTUAL PROPERTY

This solicitation affords proposers the choice of submitting complete program proposals for the award of a grant, a Sponsored Research Agreement, or other agreement as appropriate. Intel reserves the right to negotiate the final choice of agreement. The final award terms are expected to follow a public dedication model. This means that Intel and the university will jointly agree that IP developed under an award will be placed in the public domain, including publishing (and not patenting) all inventions and offering software under a permissive open source license such as BSD, MIT, or Apache 2.0. Proposed exceptions to this model must be described in the RFP and will be considered at Intel's sole discretion.

POINT OF CONTACT FOR INQUIRIES AND SUBMISSIONS

Please complete the cover sheet at the end of this RFP and include with your proposal. Proposal submissions (and related inquiries) should be directed to:

Jeff Parkhurst, Program Director, Intel (jeff.parkhurst@intel.com)

Sreenivas Subramoney, Principal Investigator, Intel (Sreenivas.Subramoney@intel.com)

David Koufaty, Principal Engineer, Intel (David.Koufaty@intel.com)

This RFP is administered by the Intel Lab's University Research & Collaboration Office. Staff overseeing this program include Gabriela Cruz Thompson and Gil Vandentop.