FoMR Webinar Script

Slide 1: **[Jim Kurose]**

Good afternoon! I'm Jim Kurose, the Assistant Director of the National Science Foundation (NSF) for Computer and Information Science and Engineering (CISE). With me here today are Yan Solihin, Tao Li, Sandip Kundu, Anindya Banerjee, and Rao Kosaraju. We're also pleased to have Matthew Haycock, Jeff Parkhurst, and Hong Wang join us from Intel via phone.

On behalf of my NSF and Intel colleagues, it is my great pleasure to welcome you to this webinar about our new, joint, NSF/Intel program solicitation on Foundational Microarchitecture Research, which we are abbreviating as "FoMR" ("F-O-M-R").

The FoMR program represents the latest in a series of joint program solicitations that NSF and Intel have published together. Our goal with these joint programs has been to catalyze advances in fundamental, pre-competitive research in emerging areas of mutual interest. These programs have built upon NSF's experience with developing and managing large, diverse research portfolios, along with Intel's history of growing research communities in specific areas.

This particular solicitation, in foundational microarchitecture, deepens our mutual commitment to support transformative research that advances computer performance through innovative computer architecture designs.

We therefore anticipate the research funded by the FoMR program will support innovative approaches to microarchitecture design, with a particular focus on improving instructions per cycle, an important component in the performance growth of future computer systems.

Research in this area is critical for providing the performance that will help advance all areas of science, technology, engineering, and mathematics. It will also contribute to advances in nearly all sectors of society, given the increasing ubiquity and pervasiveness of computing as the driver of economic competitiveness.

Before I close, I want to take a moment to acknowledge and thank Matt, Hong, and Jeff at Intel for their excellent partnership in helping to establish the FoMR program. We are looking forward to continuing to work together with them to manage this program and ensure its success.

Let me also thank my colleagues here at NSF – including Yan Solihin, Tao Li, Sandip Kundu, Anindya Banerjee, and Rao Kosaraju – for their leadership in this effort. Their vision and passion has enabled the FoMR program.

We look forward to bold new ideas from all of you – our research community – as we aim to enable performance growth for microprocessors well into the future.

Let me now turn things over to Matt Haycock, Director of Architecture and Design Research at Intel. Matt leads research into silicon architecture, circuit and design innovations to enable future capabilities and experiences across all segments of computing. Matt has a history of extremely successful research in this area, both at Intel and previously at Philips Semiconductors. Thanks very much, Matt.

[Matthew Haycock]

Good afternoon. Thank you for joining us in this webinar today. I'm Matthew Haycock, Director of Architecture and Design Research at Intel. I'm joined today by Hong Wang, Intel Fellow and Director of Intel's Microarchitecture Research Lab and by Jeff Parkhurst who drives a group of Intel's academic programs and collaborations.

We are pleased to be able to partner with the National Science Foundation in this exciting program to reinvigorate microarchitecture research within the academic community. We want to acknowledge and thank Yan Solihin and Tao Li for the outstanding partnership we have enjoyed with them during the formation of this NSF/Intel Partnership on Foundational Microarchitecture Research. They have been great partners and we look forward to this unique and important opportunity.

This partnership, FoMR, aims to combine the strengths of the NSF in sponsoring and shaping strong research in foundational areas with Intel's perspective on the current state of CPU and SOC microarchitecture research and on interesting opportunities that lie ahead in this field.

For more than 50 years, our industry has enjoyed the benefits of Moore's Law scaling. Additionally and more recently, we are seeing the results of research exploiting thread-level and data-level parallelism to increase throughput performance. Couple these trends of transistor scaling and parallelism with slowing clock frequency growth and increasing architectural designs on each process node, and with the rapid growth new areas such as machine learning, artificial intelligence, and non-volatile memory, and we see an increasing opportunity for a fresh look at innovative microarchitecture techniques to increase single thread performance and at methodologies to deliver performance growth in the future.

Intel sees the confluence of these trends producing a fertile landscape in which to reinvigorate and accelerate foundational microarchitecture research. Through this partnership with the NSF, we want to create a vibrant research ecosystem wherein academic researchers can develop and use a common set of tools with which to efficiently explore many new microarchitecture ideas and converge on a consistent set of benchmarks by which designs can be compared and contrasted. In this ecosystem, we also want to foster innovation by sharing research results and encouraging open source and publication practices, all aimed at discovering the foundational architecture and microarchitecture technologies that will fuel growth in IPC and instruction-level parallelism in the years ahead.

We invite and encourage proposals for innovation in hardware or software and that will practice open source approaches to facilitate reproducibility and reuse of research results by the broader microarchitecture research community.

Again – thank you for participating today and we look forward to the opportunities this partnership with the NSF and the academic community will provide. I will now turn the discussion over to Yan Solihin who will describe the partnership in more detail.

[Yan Solihin]

Thank you, Matt.

Slide 2: Good afternoon. I'm Yan Solihin, and one of the program directors managing NSF/Intel Partnership on Foundational Microarchitecture Research, or FoMR. Beside me is Tao Li, another program director managing FoMR. On the Intel side, we have Matt Haycock, Jeff Parkhurst, and Hong Wang. In this webcast, we'll give a brief overview of the program and some of the most important things you need to know about submitting a proposal.

Slide 3: Here is the agenda for today's presentation. we'll start with a brief overview of the FoMR program. Next, we'll cover some important aspects of the solicitation including the types of awards to be made, submission requirements, and of course the deadline. we'll cover a few points to remember, and finally we'll invite further questions from the audience and answer them with help from our colleagues.

Slide 4: As Jim Kurose mentioned, FoMR will support transformative microarchitecture research targeting improvements in instructions per cycle (IPC). This solicitation seeks microarchitecture technique innovations beyond simplistic, incremental scaling of existing microarchitectural structures.

Slide 5:
FoMR is seeking the following characteristics:

1. high IPC techniques ranging from microarchitecture to code generation;
2. "microarchitecture turbo" techniques that marshal chip resources and system memory bandwidth to accelerate sequential or single-threaded programs; and
3. techniques to support efficient compiler code generation.

Success in this area promises to provide significant performance improvements to continue the cadence promised by Moore's Law.

Slide 6:
Example areas of interest (not exhaustive) are:

- **Microarchitecture and code generation techniques** to boost branch prediction accuracy puts an upper bound on the ILP achievable by programs. There are still branches that are very hard to predict, and these offer an opportunity to push the upper bound higher. Similar situations apply to memory disambiguation and cache management: on average, these are working well, but some behavior remains hard to predict. Cache miss ratios are low for many programs but are high for some other programs; they also place an upper bound on the achievable ILP.
- **Instruction scheduling and organization of execution resources** to enable efficient, very-large instruction window processors. Improving ILP requires the instruction window to be enlarged substantially. A very-large instruction window creates challenges in complexity and lengthens critical path delay.
- **Next-generation prefetchers** with vastly higher coverage than possible today, especially in light of emerging memory technologies. Emerging memory technologies, while non-volatile, have read latencies that are often higher than dynamic random access memory (DRAM), and write latencies that are much higher than DRAM. In addition, they often have write endurance problems. Techniques to redesign prefetching so that coverage can be increased substantially are desired. Techniques to redesign prefetching so that it works well with emerging memory technologies are desired as well.
- **Cache management** techniques that work efficiently across multiple levels of cache hierarchy comprising heterogeneous memories [e.g., static random access memory (SRAM), high bandwidth memory (HBM), or traditional DRAM]. These may include cache allocation, replacement, bypassing, prefetching, etc., in the context of heterogeneous memories.
- **Machine learning and data analytics-based approach** to improving microarchitecture design and runtime tuning. Machine learning and data analytics provide new tools for improved ability to understand and optimize microarchitecture and program behavior. These tools may be utilized for runtime tuning of microarchitecture parameters so that they can be adapted to execute programs more efficiently.

- **Utilization of new microarchitecture building blocks**, such as reconfigurable logic to boost IPC. As fine-grain reconfigurable fabric [e.g., classic field-programmable gate arrays (FPGA)] and coarse-grain reconfigurable arrays (CGRA) become integrated into microprocessors, they have potential to become the foundational implementation building block akin to SRAM and the register file to enable novel microarchitecture structures for general-purpose microprocessors. For example, they can be used to construct workload-specific branch predictors or memory prefetchers that comprehend salient patterns in control flow behavior or memory access behavior unique to workload execution, in particular, cloud computing. Enabling such deployment-aware, workload-specific predictors that can be programmed and updated in-the-field is akin to a traditional microcode update in the virtualized environment.
- **Microarchitecture support for efficient compiler code generation** beyond trace generation and traditional performance counter profiling. There is a need for microarchitectures to provide higher-fidelity feedback to the compiler, so that the compiler can generate more efficient code, either for recompilation purposes or for just-in-time optimizations.
- **Criticality-oriented design and techniques.** Critical path delay limits ILP and the overall performance of computer systems. Techniques that reduce the critical path delay, and at the same time exploit slacks on non-critical paths in order to boost ILP, are desired.
- **Microprocessor and process co-optimization.** Microprocessors and processes are often designed separately. However, there are opportunities for co-design, which in turn would enhance optimization, and these should be exploited when possible and practical to improve ILP.

Slide 7:
Here are some highlights from the FoMR solicitation. Please note that while we are spotlighting key pieces of information you need to know in order to submit a proposal, this is not a substitute for reading the full solicitation. You can easily find the solicitation by using the search string N-S-F space F-o-M-R.
A proposal can request up to $500K and span up to three years.

We expect to make around six awards of up to $500K, subject to availability of funds.
The deadline for submitting proposals is January 12, 2018.

Slide 8:
Universities and two- and four-year colleges in the US can submit proposals, on behalf of their faculty.
There are no restrictions or limits on who may serve as PI.
There is a limit for the number of proposals per PI or co-PI. An investigator may participate as PI, co-PI, or senior personnel in no more than one proposals submitted in response to this solicitation.

Slide 9:
There are 4 types of supplementary documentation required:

- First is the list of project personnel;
- Second is the collaboration plan, which we will describe in more detail in just a moment;
- Next is the data management plan [this is where you should discuss any planned release of applications, tools, languages, compilers, libraries, architectures, systems, data, and so on, which we encourage]; and
- Finally, there's the postdoc mentoring plan, which is required in any proposal that includes funds for a postdoc.

Slide 10:
Each proposal with at least two investigators is required to have a collaboration plan as a separate supplementary document, which must describe the backgrounds and expertise of the PIs and how the PIs plan to collaborate. This document will be evaluated by the panelists or reviewers as part of the proposal review process. It is up to the proposers to make the argument that they provide distinct expertise in this section.

If the proposal has at least two investigators but the collaboration plan is missing, your proposal will be returned without review.

Slide 11:
While we wait for the questions to come in, we'll answer some points to remember about the FoMR program.

The first question is: How can I tell whether my proposed research is a good fit for the program?
As mentioned earlier, this solicitation is seeking proposals with the following characteristics:

- Techniques targeting high Instructions per Cycle (IPC);
- Microarchitecture ``turbo": marshalling chip resources and system bandwidth to boost single thread performance; and
- Efficient code generation targeting IPC.

Slide 12:
The next question is: Should I discuss my proposal with NSF Program Directors?
Yes, we encourage you to discuss planned proposals with Program Directors to help determine fit to the program. We ask that you refrain from scheduling separate meetings or calls with multiple Program Directors. Once submitted, the substance of proposals will not be discussed by NSF Program Directors, as this would constitute unfair competition, or the perception thereof.

Slide 13:
The next question is: Do FoMR proposals count against the CISE Core program limits on number of proposals allowed per year?
No, but there is an FoMR program limit. No one can be PI, co-PI, or senior personnel on more than one FoMR proposal.

Slide 14:
That brings us to the conclusion of the presentation. The slides and the script for this webcast, as well as an audio recording, will be available via the NSF Events web page; the URL is shown (and you may have followed it to access this webcast). On that page, you'll need to look for this webcast among the list of events. Also, the email addresses of the NSF FoMR program directors are shown, along with the CISE divisions they belong to. The email addresses of Intel contacts are also shown. If you can't determine which one to contact based on division or the core programs they manage, just contact us. Thank you all for your attention.