Solicitation No. DARPA-PA-19-04-02
Microsystems Exploration Topic (μE)
Performant Automation of Parallel Program Assembly (PAPPA)

I. Topic Description

The Defense Advanced Research Projects Agency (DARPA) is issuing a Microsystems Exploration topic (μE) inviting submissions of innovative basic or applied research concepts in the technical domain of massively parallel heterogeneous computing, parallel programming methods, self-modifying compilers, program synthesis, and predictive system modeling. This μE is issued under the Program Announcement for Microsystems Exploration, DARPA-PA-19-04. All proposals in response to the technical area described herein will be submitted to DARPA-PA-19-04 and if selected, will result in an award of an Other Transaction (OT) for prototype project not to exceed $1,000,000.

A. Introduction

With the expected tapering of transistor density and performance scaling, future hardware performance gains must be derived from massive distributed parallelism, function specialization, and/or extreme system heterogeneity. In order to make these performance gains accessible to the DoD, we need new programming methods that enable near peak hardware performance with minimal programming effort. Decades of research in high productivity computing has yielded methodologies and languages that offer either programmer productivity or performance scalability, but never both simultaneously. Current high productivity programming frameworks enable rapid prototyping, but they do not generate executables efficient enough for deployment. In contrast, performant programs suitable for massively parallel hardware require significant time and domain expertise to manually control and specify parallelization, interprocess communication, and synchronization. The PAPPA opportunity seeks novel programming approaches and compiler technology that enable scientists and application experts to efficiently compile a broad set of high-level mathematical programs to extremely parallel and heterogeneous hardware.

B. Objective/Scope

The PAPPA topic will explore tradeoffs between programming productivity, solution generality, and scalability to enable scientists and domain experts with no understanding of parallel programming and hardware architectures to create highly efficient performance portable programs. The goal of PAPPA is to explore the creation of compiler technology that improves the programming productivity of massively parallel and heterogeneous processing systems with 1,000,000 way parallelism by 10,000X. If successful, such performance portable compilers will significantly lower the barrier to deploying new algorithms and science on future state of the art programmable systems.

The majority of today's supercomputer scale high performance programs are implemented using a combination of C/C++/Fortran and Message Passing Interface (MPI) which requires the programmer to explicitly control parallelization and communication. MPI style programming has demonstrated scalability to millions of cores, but comes with unacceptably high software expertise barriers, long development times, and high development costs. Attempts to create programming

frameworks that automate parallelization have resulted in moderate speedups for shared memory systems, but these methods are not expected to scale to thousands or millions of nodes. At the other end of the programming productivity spectrum, frameworks like Python, R, and Octave have demonstrated extremely good programming productivity, but they have not demonstrated general purpose scalability on massively parallel systems.

In 2002, to address these challenges, DARPA stood up the High Productivity Computing Systems program (HPCS).[1] The HPCS program objective was to create disruptive enhancement in program performance, programmability, and portability for High Performance Computing (HPC). While the program resulted in significant advances in hardware architectures and parallel programming methods, the grand challenge of simultaneous productivity and performance has remained elusive.

A number of recent computer science trends suggest that this elusive grand challenge may now be within reach. This includes:

- Success of domain specific languages that demonstrate simultaneous high performance and productivity (e.g. TensorFlow, PyTorch, etc.).
- Advances in the machine learning field that could be applied to speed up automated system modeling, characterization, and search for an optimal solution across the extremely large search space of mapping, scheduling, and distribution possibilities.
- Proliferation of open source compilers and standardized intermediate representations (e.g. LLVM) allowing rapid exploration and end-user programming model testing.

To enable productive high performance programming going forward, it is imperative that all low level tasks including task partitioning, aggregation, mapping, scheduling, resource allocation, memory management, inter process communication, synchronization, and reliability be removed as burdens from the programmer. To enable scalable and performance portable generation of executables from a highly productive programming framework, it seems likely that a completely new approach is needed. PAPPA hypothesizes that many of the challenges associated with the high performance programing of massively parallel heterogeneous systems can be overcome by accurately modeling and predicting performance of all key components within the system and applying global cost optimization to successively lower and optimize appropriately constrained domain specific programs to generate highly efficient executable code.

The PAPPA Figure of Merit (FOM) can be defined as the following relationship:

HPP α (PG) x (PE) x (PP)

Where HPP stands for “High Performance Productivity”:

- **PG (Productivity Gain):** Ratio of the total development time (hours) of a state of the art implementation compared to the proposed approach for a given application. This factor

[1] Jack Dongarra, Robert Graybill, William Harrod et al. “DARPA's HPCS Program: History, Models, Tools, Languages”, Advances in Computers, 2008

measures the productivity gain of the programming framework.

- **PE (Peak Efficiency):** Ratio of the execution time (hours) of a state of the art implementation compared to the proposed approach for a given application running on identical hardware. This factor measures the performance gain of the programming framework.
- **PP (Performance Portability)**: The ratio of the maximum and minimum execution times (hours) on the state of the art implementation compared to the proposed approach for a given application tested on a set of diverse hardware platforms. [PP = (MaxOld/MinOld) / (MaxNew/MinNew).] This factor measures the compiler's ability to effectively target multiple hardware platforms without programmer involvement.

The proposed solutions should include demonstration plans for thousand node systems with analysis showing scalability to million way parallelism. Proposals should clearly explain how their HPP (and its constituent components) are calculated and should describe the baseline state of the art application(s) and hardware systems against which the HPP is benchmarked. A minimum of three separate high performance computing platforms should be targeted for benchmarking, with representation from two unique central processing unit (CPU) architectures and two unique graphics processing unit (GPU) architectures.

The programming frameworks and compiler proposed should be applicable to a broad set of mathematical programs. For demonstration purposes, the following two application domains are listed as examples, and are of particular interest; however, proposers may propose other DoD-relevant applications that would demonstrate similar capabilities and benefits.

Application Domain 1: Physical Simulation

A wide range of HPC applications can be described as physical simulations. Examples include computational simulation of fluid dynamics, weather, molecular dynamics, and particle physics. These applications are generally characterized by massive data sets, high numerical fidelity, and soft deadline latency requirements. There are numerous examples of successful application of Single Program Multiple Data programming models to this domain.

Application Domain 2: Real Time Processing

A second set of important HPC applications can be found in edge computing platforms such as radar and wireless communication systems. Applications in this domain are characterized by noisy real world sampled data, extreme throughput requirements, and hard deadline latency requirements. There are numerous examples of successful application of Dataflow programming models to this application domain.

The goal of the research should be an HPP figure of merit improvement factor of no less than 10,000X compared to existing high performance programming methods.

C. Structure

Proposals submitted to DARPA-PA-19-04 in response to this μE topic must be UNCLASSIFIED and must address two independent and sequential project phases: a Phase 1 Feasibility Study

(base) and a Phase 2 Proof of Concept (option). The periods of performance for these phases are 6 months for the Phase 1 base effort and 12 months for the Phase 2 option effort. Combined Phase 1 base and Phase 2 option efforts for this μE topic should not exceed 18 months. The Phase 1 (base) award value should not exceed $250,000. The Phase 2 (option) award value should not exceed $750,000. The total award value for the combined Phase 1 and Phase 2 is limited to $1,000,000.

Phase 1 studies will be evaluated to determine the feasibility of the approach and whether to exercise the option for Phase 2.

D. Schedule/Milestones

Proposers must address the following Research Project Objectives, metrics, and deliverables, along with fixed payable milestones in their proposals. A Schedule of Milestones and Payments excel spreadsheet has been provided as an attachment to this PAPPA solicitation. Proposers are required to submit this attachment with their proposal submission, with priced milestone payments consistent across the Task Description Document (TDD) and the Vol. 2 – Price Volume. The task structure must be consistent across the proposed schedule, Task Description Document (TDD), and the Vol. 2 - Price Volume. If selected for award negotiation, the fixed payable milestones will be directly incorporated into Attachment 2 of the OT agreement ("Schedule of Milestones and Payments"). Please see the sample OT for Prototype provided as an attachment to DARPA-PA-19-04.

For planning and budgetary purposes, proposers should assume a program start date that is 90 days from μE topic announcement on fbo.gov. Schedules will be synchronized across performers, as required, and monitored/revised as necessary throughout the program.

All proposals must include the following meetings and travel in the proposed schedule and costs:

- To foster collaboration between teams and disseminate program developments, a two-day Principal Investigator (PI) meeting will be held approximately every six months, with locations split between the East and West Coasts of the United States. For budgeting purposes, plan for three two-day meetings over the course of 18 months: two meetings in the Washington, D.C. area and one meeting in the San Francisco, CA area.
- Regular teleconference meetings will be scheduled with the Government team for progress reporting as well as problem identification and mitigation. Proposers should also anticipate at least one site visit per phase by the DARPA Program Manager during which they will have the opportunity to demonstrate progress towards agreed-upon milestones.

Phase 1:

Objective: Develop a compiler architecture, distributed programming framework, intermediate representation, and system modeling approach compatible with high productivity/high performance compilation for extreme parallelism and heterogeneity.

Milestones:

- Month 1: Provide a comprehensive description of the proposed compiler approach with in depth support from literature. Include an overview and comparison of the current state of

the art pertaining to the proposed approach and an initial description of the applications, hardware systems, and benchmarks that will be used for compiler comparison. Describe how the proposed compiler architecture will meet the required performance and productivity goals of the program.

- Month 3: Provide an interim report describing the approach, historical application comparison data, and estimated performance and scalability analysis for the chosen compiler approach.
- Month 5: Submit Phase 1 Final report in preparation for Phase 2 option. The report should summarize the chosen approach for phase 2 tasks including a detailed description of the prototype compiler architecture and kernel level proof points demonstrating productivity and performance portability consistence with the PAPPA FOM goals.

Phase 2:

Objective: Develop a prototype compiler that demonstrates the feasibility of achieving the PAPPA opportunity HPP FOM goal for a broad range of hardware targets and real world applications.

Milestones:

- Month 9: Report describing initial results of a prototype compiler implementation.
- Month 12: Live demonstration of end to end prototype compiler working demonstrating successful compilation of high level code into executable code.
- Month 15: Live demonstration of a compiler framework to illustrate scalability, productivity, and performance portability across multiple CPU and GPU hardware platforms.
- Month 18: Demonstration of functional compiler framework. Provide a Phase 2 report that documents the detailed implementation of the compiler architecture and FOM metrics tracked across a range of applications. Provide comprehensive recommendations for future paths for extending the compiler framework to more applications and more hardware platforms.

E. Deliverables

Performers will be expected to provide at a minimum the following deliverables:

Phase 1:

- Distributed computing language specification
- Intermediate representation specification (IR)
- Component, network, and memory system modeling specification
- Final Phase 1 report summarizing compiler approach, software architecture, applications tested, data sets, and prototype results.
- All other reports and data as required by the individual Phase 1 milestones.

Phase 2:

- Demonstration of multi-target compiler software for distributed high performance computing

- Updated distributed computing language specification
- Updated intermediate representation (IR) specification (if applicable)
- Updated component and network modeling specification
- Compiler software documentation
- Prototype compiler
- Final Phase 2 report documenting final compiler approach, software architecture, analysis of compiler metrics across application domains tested.

Software created in the program is expected to be readily adopted and leveraged by the broader high performance computing research community, so intellectual property rights asserted by proposers are strongly encouraged to be aligned with open source licenses such as Apache 2.0, Massachusetts Institute of Technology (MIT), Berkeley Software Distribution, and GNU public licenses. If a proposed approach includes proprietary software or technical data as a component of the approach, the proposer is expected to provide 1) clear justification for the need for the proposed software, and 2) a description of how the PAPPA program goals will be met with use of the proprietary model.

II. Award Information

Selected proposals that are successfully negotiated will result in award of an OT for prototype project. See Section 3 of DARPA-PA-19-04 for information on awards that may result from proposals submitted in response to this notice.

Proposers must review the model OT for Prototype agreement provided as an attachment to DARPA-PA-19-04 prior to submitting a proposal. DARPA has provided the model OT in order to expedite the negotiation and award process and ensure DARPA achieves the goal of Microsystems Exploration, which is to enable DARPA to initiate a new investment in less than 90 days from each μE topic announcement. The model OT is representative of the terms and conditions that DARPA intends to award for all Microsystems Exploration Awards. The task description document, schedule of milestones and payments, and data rights assertions requested under Volumes 1, 2, and 3 will be included as attachments to the OT agreement upon negotiation and award.

Proposers may suggest edits to the model OT for consideration by DARPA and provide a copy of the model OT with track changes as part of their proposal package. Suggested edits may or may not be accepted by DARPA. In order to ensure that DARPA achieves the Microsystems Exploration goal of award within 90 days from the posting date of the μE topic announcement, DARPA reserves the right to cease negotiations if the parties fail to reach agreement on OT award terms and conditions within this time period. If edits to the model OT are not provided as part of the proposal package, DARPA assumes that the proposer has reviewed and accepted the award terms and conditions to which they may have to adhere and the sample OT agreement provided as an attachment, indicating agreement (in principle) with the listed terms and conditions applicable to the specific award instrument.

III. Eligibility

See Section 4 of DARPA-PA-19-04 for information on who may be eligible to respond to this

notice.

IV. μE Topic Responses

Responses to this μE topic must be submitted as full proposals to DARPA-PA-19-04 as described therein. All proposals must be unclassified.

A. Proposal Content and Format

All proposals submitted in response to this notice must comply with the content and format instructions in Section 5 of DARPA-PA-19-04. All proposals must use the templates provided as Attachments to the PA and follow the instructions therein.

Information not explicitly requested in DARPA-PA-19-04, its Attachments, or this notice may not be evaluated.

B. Proposal Submission Instructions

See Section 5 of DARPA-PA-19-04 for proposal submission instructions.

C. Proposal Due Date and Time

Proposals in response to this notice are due no later than 4:00 PM on October 3, 2019. Full proposal packages as described in Section 5 of DARPA-PA-19-04 must be submitted per the instructions outlined therein and received by DARPA no later than the above time and date. Proposals received after this time and date may not be reviewed.

Proposers are warned that the proposal deadline outlined herein is in Eastern Time and will be strictly enforced. When planning a response to this notice, proposers should take into account that some parts of the submission process may take from one business day to one month to complete.

V. Proposal Evaluation and Selection

Proposals will be evaluated and selected in accordance with Section 6 of DARPA-PA-19-04. Proposers will be notified of the results of this process as described in Section 7.1 of DARPA-PA-19-04.

VI. Administrative and National Policy Requirements

Section 7.2 of DARPA-PA-19-04 provides information on Administrative and National Policy Requirements that may be applicable for proposal submission as well as performance under an award.

VII. Point of Contact Information

Andreas Olofsson, Program Manager, DARPA/MTO, PAPPA@darpa.mil

VIII. Frequently Asked Questions (FAQs)

All technical, contractual, and administrative questions regarding this notice must be emailed to PAPPA@darpa.mil. Emails sent directly to the Program Manager or any other address may result in delayed or no response.

All questions must be in English and must include name, email address, and the telephone number of a point of contact. DARPA will attempt to answer questions publically in a timely manner; however, questions submitted within 7 days of the proposal due date listed herein may not be answered.

DARPA will post an FAQ list under the μE topic on the DARPA/MTO Opportunities page at (http://www.darpa.mil/work-with-us/opportunities). The list will be updated on an ongoing basis until one week prior to the proposal due date. In addition to the FAQ specific to this notice, proposers should also review the Program Announcement for Microsystems Exploration General FAQ list on the DARPA/MTO Opportunities page under the Program Announcement (DARPA-PA-19-04).