Second Workshop about the SparCity framework and the modeling tools

Deliverable No:	D7.9
Deliverable Title:	Second Workshop about the SparCity framework and the modeling tools
Deliverable Publish Date:	30 November 2023
Project Title:	SPARCITY: An Optimization and Co-design Framework for Sparse Computation
Call ID:	H2020-JTI-EuroHPC-2019-1
Project No:	956213
Project Duration:	36 months
Project Start Date:	1 April 2021
Contact:	sparcity-project-group@ku.edu.tr

List of partners:

Participant no.	Participant organisation name	Short name	Country
1 (Coordinator)	Koç University	KU	Turkey
2	Sabancı University	SU	Turkey
3	Simula Research Laboratory AS	Simula	Norway
4	Instituto de Engenharia de Sistemas e Computadores, Investigação e Desenvolvimento em Lisboa	INESC-ID	Portugal
5	Ludwig-Maximilians-Universität München	LMU	Germany
6	Graphcore AS*	Graphcore	Norway

*Until M21

CONTENTS

1 INTRODUCTION

The SparCity project is funded by EuroHPC JU (the European High Performance Computing Joint Undertaking) under the 2019 call of Extreme Scale Computing and Data Driven Technologies for research and innovation actions. SparCity aims to create a supercomputing framework that will provide efficient algorithms and coherent tools specifically designed for maximizing the performance and energy efficiency of sparse computations on emerging High Performance Computing (HPC) systems, while also opening up new usage areas for sparse computations in data analytics and deep learning.

Sparse computations are commonly found at the heart of many important applications, but at the same time it is extremely challenging to achieve high performance when performing the sparse computations. SparCity delivers a coherent collection of innovative algorithms and tools for enabling high efficiency of sparse computations on emerging hardware platforms. More specifically, the objectives of the project are:

- to develop a comprehensive application and data characterization mechanism for sparse computation based on the state-of-the-art analytical and machine-learning-based performance and energy models,

- to develop advanced node-level static and dynamic code optimizations designed for massive and heterogeneous parallel architectures with complex memory hierarchy for sparse computation,

- to devise topology-aware partitioning algorithms and communication optimizations to boost the efficiency of system-level parallelism,

- to create digital SuperTwins of supercomputers to evaluate and simulate what-if hardware scenarios,

- to demonstrate the effectiveness and usability of the SparCity framework by enhancing the computing scale and energy efficiency of challenging real-life applications.

- to deliver a robust, well-supported and documented SparCity framework into the hands of computational scientists, data analysts, and deep learning end-users from industry and academia.

1.1 OBJECTIVES OF THIS DELIVERABLE

The goal of this deliverable is to describe the proceedings of the second workshop organized to disseminate the results of SparCity to endusers and to gather feedback on the usability of the developed tools and methodologies.

The workshop was proposed to the Euro-Par 2023 organizers as a workshop co-located with that conference. The proposal was accepted and the workshop was held as the *Workshop on Tools for Data Locality, Power and Performance (TDLPP 2023)* on August 29, 2023 in Limassol, Cyprus. A workshop webpage was set up at `https://www.tdlpp-workshop.org/` and the paper submission and review process was set up using EasyChair.

There were about 30 participants present at the workshop which proved to be a venue for the exchange of of cutting-edge research, fostering collaboration and networking among experts in the field. Interactive discussions and hands-on activities contributed to a vibrant atmosphere where participants could share innovative ideas and establish new connections.

1.2 WORKSHOP OVERVIEW

The goal of the TDLPP workshop was to provide a venue for developers and users of tools that address the important topic of memory access optimization. While hardware continues to evolve and high-bandwidth memory becomes available in accelerators and mainstream CPUs, the gap between compute capability (in terms of arithmetic operations per second) and the speed of memory (in terms of access latency or amount of bytes transferred) continues to widen. Tools are thus needed to help developers understand the behavior of their codes to support them with optimizing and modeling their applications. This is especially true in application areas that involve sparse matrices, tensors, or graphs.

In 2023 we organized the inaugural edition of the TDLPP workshop co-located with the Euro-Par 2023 conference in Limassol, Cyprus. We received seven submissions and an international workshop committee provided at least three reviews for each paper. Five papers were selected for presentation in person at the workshop covering topics describing general advances in performance tool technology as well as approaches specifically addressing sparse computing uses cases. The latter was a special focus area of the workshop motivated by the Euro-HPC project SparCity (www.sparcity.eu) from which the idea for the TDLPP workshop developed.

In addition to the main session for presenting workshop papers, we continued the workshop with an interactive tool demo session to allow contributors more time showcasing their software to interested attendees. The tool demo session featured two contributions. The first contribution was provided by Fatih Taşyaran (Sabancı University), Osman Yasal (Koç University), José António Carvalho Freire Morgado (INESC-ID) showcasing "SuperTwin: A Digital Twin for HPC Machines". The second contribution was provided by Olaf Krzikalla who provided an interactive demo of his tool for "Analyzing One-Sided Communication Using Memory Access Diagrams". The tool demo session was well received and should be considered again for future iterations of the workshop.

1.3 ORGANIZATION COMMITTEE

Karl Fürlinger	LMU Munich, Germany
Didem Unat	Koç University, Turkey

1.4 PROGRAM COMMITTEE

The program committee for the workshop was recruited internationally both from within the SparCity project and from external experts.

Xing Cai	Simula Research Laboratory, Norway
Karl Fürlinger	LMU Munich, Germany
Aleksandar Ilic	INESC-ID, Portugal
Humayun Kabir	Microsoft, USA
Kamer Kaya	Sabancı University, Turkey
Johannes Langguth	Simula Research Laboratory, Norway
Leonel Sousa	INESC-ID, Portugal
Nathan Tallent	Pacific Northwest National Laboratory, USA
Miwako Tsuji	RIKEN, Japan
Didem Unat	Koç University, Turkey
Josef Weidendorfer	Leibniz Supercomputing Centre, Germany

1.5 CONTRIBUTIONS

Here we provide a list of all contributions to the workshop. This includes papers presented at the workshop as well as tool demonstrations provided by the workshop participants.

- Paper: *Leveraging HPC Profiling & Tracing Tools to Understand the Performance of Particle-in-Cell Monte Carlo Simulations.* Jeremy Williams, David Tskhakaya, Stefan Costea, Ivy Peng, Marta Garcia-Gasulla and Stefano Markidis

- Paper: *Analyzing One-Sided Communication Using Memory Access Diagrams.* Olaf Krzikalla, Arne Rempke and Ralph Mueller-Pfefferkorn

- Paper: *Enhancing Performance Monitoring in C/C++ Programs with EDPM: A Domain-Specific Language for Performance Monitoring.* David Weisskopf Holmqvist and Suejb Memeti

- Paper: *Performance Prediction for Sparse Matrix Vector Multiplication using Structure-dependent Features.* Konstantin Pogorelov, James Trotter and Johannes Langguth

- Paper: *Sparse-aware CARM: Rooflining locality of sparse computations.* Afonso Coutinho, Diogo Marques, Leonel Sousa and Aleksandar Ilic

- Tool Demo: *SuperTwin: A Digital Twin for HPC Machines.* Fatih Taşyaran (Sabancı University), Osman Yasal (Koç University), José António Carvalho Freire Morgado (INESC-ID)

- Tool Demo: *Analyzing One-Sided Communication Using Memory Access Diagrams.* Olaf Krzikalla