ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Institut für Integrierte Systeme
Integrated Systems Laboratory

TASK ASSIGNMENT FOR A
SEMESTER THESIS AT THE DEPARTMENT OF
INFORMATION TECHNOLOGY AND ELECTRICAL ENGINEERING

SPRING SEMESTER 2024

Diego de los Santos

Implementing an OpenMP runtime for MemPool in LLVM

March 18, 2024

Advisors: Samuel Riedel, ETZ J69.2, Tel. +41 44 632 65 69, sriedel@iis.ee.ethz.ch
Sergio Mazzola, ETZ J76.2, Tel. +41 44 632 81 49, smazzola@iis.ee.ethz.ch

Professor: Prof. Dr. L. Benini

Handout: March 18, 2024
Due: July 01, 2024

1 Introduction

Striving for high image quality, even on mobile devices, has led to an explosion in the pixel count of smartphone cameras over the last decade [1]. These image sensors, boasting tens of millions of pixels, create a massive amount of data to be processed on a tight power envelope as quickly as possible. Computational photography, computer vision, augmented reality, and machine learning are only a few of the possible applications. Highly specialized image signal processors (ISPs) harness the features of these workloads, which are highly parallelizable, to meet timing and power constraints. Google offers a prominent example: the Pixel Visual Core, employed for image processing in Google smartphones [2].

At ETH, we are developing our own manycore system that can be used as an ISP called MemPool [3]. It boasts 256 area-optimized 32-bit Snitch [4]. Snitch, developed at ETH as well, implements the RISC-V instruction set architecture (ISA), which is an open ISA targeting modularity and scalability [5]. Despite its size, MemPool gives all 256 cores low-latency access to the shared L1 memory, with a maximum latency of only five cycles when no contention occurs. This implements efficient communication among all cores, making MemPool suitable for various workload domains and easy to program.

A common framework for parallel programming is OpenMP [6]. It implements a task-based or fork-join approach to parallelization through annotations in the C code. The compiler will generate routines and runtime calls to distribute the parallel execution across the specific platform correctly. To this end, each system supporting OpenMP needs to implement a corresponding OpenMP Runtime. Specifically, there are two major runtimes, the GOMP and KMP runtime for GCC and LLVM, respectively. MemPool has minimal support for the GOMP runtime. However, since LLVM has become the standard compiler infrastructure for our research group, a KMP runtime to support LLVM is crucial.

Goal The goal of this thesis is to implement the KMP runtime for MemPool. To this end, the student will first identify the most important OpenMP functions for MemPool and then implement the runtime in conjunction with functional tests for verification. For benchmarking, the student will implement and evaluate digital signal processing (DSP) kernels and potentially full applications from the image processing or communications domain.

2 Milestones

The following are the milestones that we expect to achieve throughout the project:

- Become familiar with MemPool and the OpenMP runtime.

- Investigate the most useful features of the OpenMP standard.
- Implement the key features of the KMP runtime for MemPool and functionally verify them.
- Benchmark the newly implemented runtime on several DSP kernels.

2.1 Stretch Goals

Should the above milestones be reached earlier than expected and the student is motivated to do further work, we propose the following stretch goals to aim for:

- Advance the KMP implementation by implementing features beyond the basic set determined during the initial project phase.
- Build a full demonstrator application using OpenMP. The application can be chosen during the project, but examples include an HDR image processing pipeline, a ray tracing algorithm, or a 5G communication pipeline.
- Tune your benchmarks or demonstrator application for Heartstream, a 64-core chip implementing the latest MemPool architecture.

3 Project Realization

3.1 Time Schedule

The time schedule presented in Table 1 is merely a proposition; it is primarily intended as a reference and an estimation of the time required for each required step.

Project phase	**Time estimate**
Familiarization with MemPool and OpenMP to determine key features	2 weeks
Implementation and verification of key features	4 weeks
Evaluation of your work and implementation of DSP kernels in OpenMP	4 weeks
Write report	2 weeks
Prepare presentation	2 weeks

Table 1: Proposed time schedule and investment

3.2 Meetings

Weekly meetings will be held between the student and the assistants. The exact time and location of these meetings will be determined within the first week of the project in order to fit the student's and the assistants' schedule. These meetings will be used to evaluate the status and progress of the project. Beside these regular meetings, additional meetings can be organized to address urgent issues as well.

3.3 Weekly Reports

The student is advised, but not required, to write a weekly report at the end of each week and to send it to his advisors. The idea of the weekly report is to briefly summarize the work, progress and any findings made during the week, to plan the actions for the next week, and to bring up open questions and points. The weekly report is also an important means for the student to get a goal-oriented attitude to work.

3.4 Coding Guidelines

HDL Code Style Adapting a consistent code style is one of the most important steps in order to make your code easy to understand. If signals, processes, and modules are always named consistently, any inconsistency can be detected more easily. Moreover, if a design group shares the same naming and formatting conventions, all members immediately *feel at home* with each other's code. At IIS, we use lowRISC's style guide for SystemVerilog HDL: `https://github.com/lowRISC/style-guides/`.

Software Code Style We generally suggest that you use style guides or code formatters provided by the language's developers or community. For example, we recommend LLVM's or Google's code styles with `clang-format` for C/C++, PEP-8 and `pylint` for Python, and the official style guide with `rustfmt` for Rust.

Version Control Even in the context of a student project, keeping a precise history of changes is *essential* to a maintainable codebase. You may also need to collaborate with others, adopt their changes to existing code, or work on different versions of your code concurrently. For all of these purposes, we heavily use *Git* as a version control system at IIS. If you have no previous experience with Git, we *strongly* advise you to familiarize yourself with the basic Git workflow before you start your project.

3.5 Report

Documentation is an important and often overlooked aspect of engineering. A final report has to be completed within this project.

The common language of engineering is de facto English. Therefore, the final report of the work is preferred to be written in English.

Any form of word processing software is allowed for writing the reports, nevertheless the use of LaTeX with Inkscape or any other vector drawing software (for block diagrams) is strongly encouraged by the IIS staff.

If you write the report in LaTeX, we offer an instructive, ready-to-use template, which can be forked from the Git repository at `https://iis-git.ee.ethz.ch/akurth/iisreport`.

Final Report The final report has to be presented at the end of the project and a digital copy needs to be handed in and remain property of the IIS. Note that this task description is part of your report and has to be attached to your final report.

3.6 Presentation

There will be a presentation (15 min presentation and 5 min Q&A) at the end of this project in order to present your results to a wider audience. The exact date will be determined towards the end of the work.

4 Deliverables

In order to complete the project successfully, the following deliverables have to be submitted at the end of the work:

- Final report incl. presentation slides
- Source code and documentation for all developed software and hardware
- Testsuites (software) and testbenches (hardware)
- Synthesis and implementation scripts, results, and reports

References

[1] S. Skafisk, *This is How Smartphone Cameras Have Improved Over Time*, 2017 (accessed August 18, 2020). [Online]. Available: https://petapixel.com/2017/06/16/smartphone-cameras-improved-time/

[2] J. Redgrave, A. Meixner, N. Goulding-Hotta, A. Vasilyev, and O. Shacham, "Pixel Visual Core: Google's fully programmable image, vision and AI processor for mobile devices," in *2018 IEEE Hot Chips 30 Symposium (HC30)*. Cupertino, US: IEEE Technical Committee on Microprocessors and Microcomputers, Aug. 2018.

[3] S. Riedel, M. Cavalcante, R. Andri, and L. Benini, "MemPool: A scalable manycore architecture with a low-latency shared L1 memory," *IEEE Transactions on Computers*, vol. 72, no. 12, pp. 3561–3575, 2023.

[4] F. Zaruba, F. Schuiki, T. Hoefler, and L. Benini, "Snitch: A tiny pseudo dual-issue processor for area and energy efficient execution of floating-point intensive workloads," *IEEE Trans. Comput.*, vol. 70, no. 11, pp. 1845–1860, Feb. 2021.

[5] A. Waterman, Y. Lee, D. A. Patterson, K. Asanovic, A. Waterman, Y. Lee, and D. Patterson, "The RISC-V instruction set manual," 2014. [Online]. Available: https://github.com/riscv/riscv-isa-manual

[6] O. A. R. Board, "OpenMP Application Programming Interface," 2021. [Online]. Available: https://www.openmp.org/wp-content/uploads/OpenMP-API-Specification-5-2.pdf

Zurich, March 18, 2024 Prof. Dr. L. Benini