LPGPU2 Report Summary

Project ID: 688759
Funded under: H2020-EU.2.1.1.

Periodic Reporting for period 2 - LPGPU2 (Low-Power Parallel Computing on GPUs 2)

Reporting period: 2017-04-01 to 2018-09-30

Summary of the context and overall objectives of the project

"Consumers today expect to be able to carry a supercomputer around with them, with detailed graphical displays, that are easy to use and last for more than a day on a single battery charge. To deliver on this capability requires very power efficient graphics processors ("GPUs") and associated software. Smart power management algorithms such as dynamic voltage frequency scaling (DVFS) are used as a mitigation for high power consumption but this usually results in a less compelling user experience as the CPU and GPU are clocked down to conserve power resulting in less raw processing capacity. The strict power limitations means that these demands cannot be met through hardware improvements alone, the software must better exploit the available resources. Unfortunately, programmers are hindered when creating low-power GPU software by the capabilities of current performance analysis tools. As software becomes more complex, it becomes increasingly unmanageable for programmers to optimize the software for low-power devices.

The main goal of the LPGPU2 project was to help developers create software for low-power GPUs by providing a complete performance and power analysis framework addressing the power problem from different angles: by defining data collection standards, reliably measuring and estimating power consumption and developing a tool suite that provides rich visualizations and insights to software developers. The LPGPU2 tool suite was built on top of the open source project CodeXL ([https://gpuopen.com/compute-product/codexl](https://gpuopen.com/compute-product/codexl)) and it consists of four main elements: data capture, data visualization, data analysis, and power estimation and measurements.

The main objectives of this project and the corresponding conclusions are:

1. To help programmers to improve the energy efficiency of compute and graphics applications for existing and emerging APIs. The LPGPU2 tool chain is equipped with a smart Feedback Engine aimed at making optimizations simple by providing insightful guidance to the user on how to improve performance and power consumption. The LPGPU2 tool suite has been validated using applications based on four existing and emerging APIs (OpenCL, SYCL, OpenGL, and Vulkan) that contain demanding graphics and compute parts.

2. To enable programmers to be able to write their software once and run it on a variety of different low-power GPUs. The project has set forward a standard interface for data collection. Establishing standard interfaces enhances also the portability of the applications across multiple platforms and standards. Moreover, the LPGPU2 tool is equipped with analysis modes able to support optimizations for four standards of the Khronos group; obviously, the use of open standards is also contributing to the portability objective.
3. To increase the productivity in GPU software development. In order to achieve this objective, the consortium decided to release the tool suite as open-source. In this way, even SMEs with limited resources have access to a sophisticated tool and enjoy the benefits of larger or more financially capable companies.

4. To reduce the hardware, software, and device driver design and development cycles of mobile GPUs. The LPGPU2 project offers a vertical toolchain. The term vertical means that the tool is able to gather (via a standardized interface) information from the GPU hardware, GPU driver, API, and the application levels and visualize this information in a seamless fashion. As such, the toolchain can be considered as a central point of reference. Consequently, it can be used as a tool to facilitate the communication between different design teams reducing in this way the long development cycles of mobile GPUs.

5. To bring technologies to market in a commercializable form, including productizing and commercializing the technologies developed in previous LPGPU (FP7 STREP) project. This includes i) bringing the SYCL standard into real-world AI applications generating commercial int"

Work performed from the beginning of the project to the end of the period covered by the report and main results achieved so far

Among others, regarding tools and power modelling, TU Berlin in partnership with Think Silicon have developed counter-based power models that can be calibrated to different hardware platforms. TU Berlin has also developed a flexible power measurement testbed, which was used for calibrating the power models and has the potential of being commercially exploited. Codeplay ported CodeXL, an open-source profiler, to be the basis for the project’s profiling and visualization tool. Samsung has proven the interception and performance counter collection through an API and a post-processing hosting environment.

On the applications side, Samsung has also developed a range of applications showcasing font rendering, augmented reality as well as virtual reality. These will be further optimized using the LPGPU2 tool suite and help improve Samsung’s mobile graphics platform, which is used by millions of people worldwide. Think Silicon has developed a set of Image Signal Processing (ISP) applications using Vulkan and the NemaGFX API. An FPGA prototype has been implemented and the NemaGFX version of the ISP algorithms has been demonstrated at industrial exhibitions. Spin Digital has developed a complete media player using its H.265 codec and a new high-performance video rendering engine that uses the latest graphics APIs (Vulkan, DX12) and allows for the creation of next generation media playback applications (Ultra-HD support, HDR, etc). These were demonstrated at the world’s largest media industry exhibitions: NAB (Las Vegas), IBC (Amsterdam), and InterBEE (Tokyo). Codeplay has ported the TensorFlow machine learning framework to OpenCL via SYCL so that the most-used AI framework in the world can run on any energy-efficient AI accelerator that supports OpenCL.

Progress beyond the state of the art and expected potential impact (including the socio-economic impact and the wider societal implications of the project so far)

The project will impact any part of European industry involved in visual applications software on portable devices. In particular, it will support European software developers working on video games and apps for smartphones, by enabling them to analyse the power consumption of their applications and provide ways to reduce the battery usage. By driving down the power consumption of mobile devices when running highly-visual software, we further enable the widespread adoption and ease-of-use of smartphones impacting society at large.

Related information