Theme: New Computing

- Sub theme: Reconfigurable AI compiler and design space exploration

With the growing diversity of AI accelerators, the significance of corresponding compiler techniques has been amplified to achieve high performance and utilization of these accelerators. It is evident that developing separate compilers for each individual accelerator would entail a significant amount of redundant effort. Moreover, such an approach hinders interoperability, as the software for each accelerator is developed in isolation without considering its relation to other accelerators. The key motivation of this research proposal stems from the realization that the optimization problem associated with deep neural networks (DNNs) exhibits a level of generality applicable to diverse scenarios, ranging from mobile devices to large-scale server clusters.
Also, the optimization problem for DNNs can be viewed as the optimization problem of a static dataflow graph, encompassing parallelization, mapping, scheduling, memory allocation, and so on. For more aggressive optimization of these problem, there is a notable surge in the exploration of deep learning-based scheduling optimization techniques, such as reinforcement learning and graph neural networks. Therefore, this research proposal aims to research AI-based compiler optimization techniques, and design space exploration methodology for HW/SW co-optimization.
In summary, we are highly interested in (but not limited to) the following list of topics.

1. Configurable design of AI compiler
2. Hardware description language for the reusable design of AI compiler
3. AI-based compiler optimization techniques (eg. Reinforcement learning, …)
4. A fast performance evaluation method for design space exploration
5. Hardware-software co-design methodology

※ The participants are also encouraged to propose new ideas outside the topics listed above.

※ Funding: Up to USD 150,000 per year