



Offloading Computation to a GPU with OpenMP

Kamer Kaya, Sabancı University

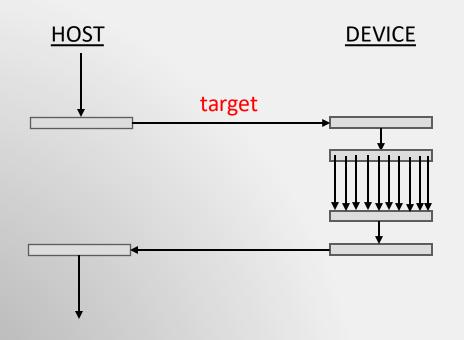


Offloading to GPU – A first encounter



Offloading to GPU - Step by step

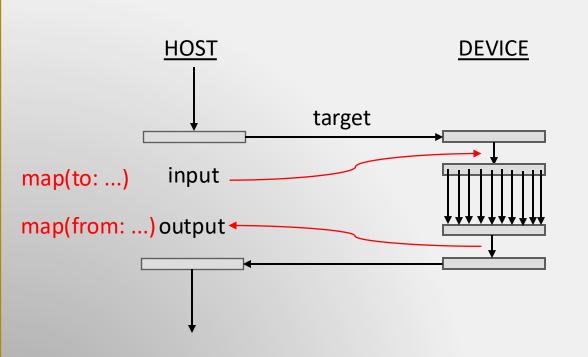




- When the host thread encounters a target region, it creates a target task for offloading to the device.
- Host temporarily suspends its own execution. The device generates an initial task, which is executed by its initial thread.
- Once the task on the accelerator is completed, the host thread resumes its previously suspended work.

Offloading to GPU - Step by step

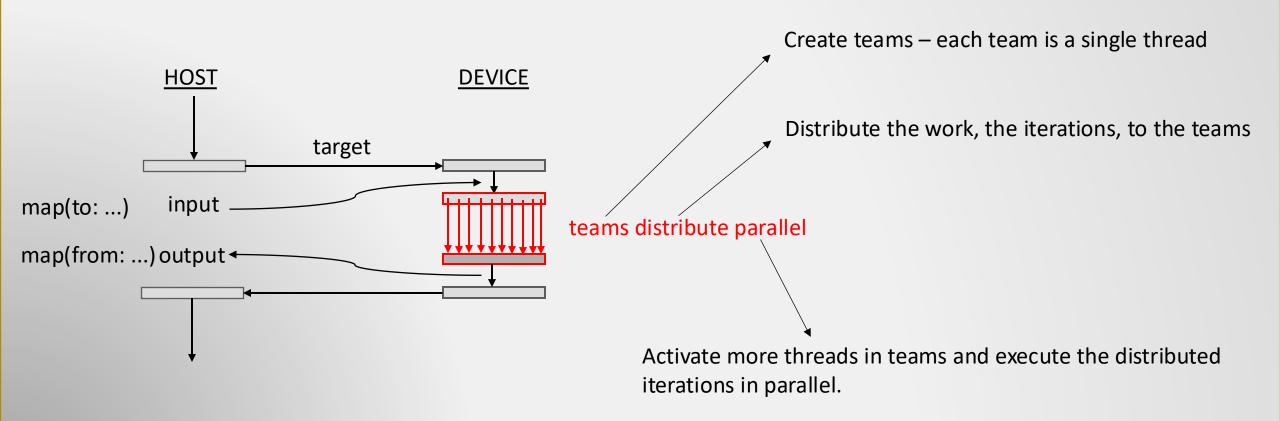




- When data from the host's environment is mapped to the device, a corresponding data is created within the device data environment.
- Throughout the execution, data are dynamically mapped and unmapped, with changes reflecting this process.
- The original host data and its device counterpart may or may not occupy the same physical memory location.

Offloading to GPU - Step by step





Example: Pi via OpenMP Offloading



Let's go over the OpenMP-to-GPU offloading code.



Thanks



This project has received funding from the European High-Performance Computing Joint Undertaking (JU) under grant agreement No 101101903. The JU receives support from the Digital Europe Programme and Germany, Bulgaria, Austria, Croatia, Cyprus, Czech Republic, Denmark, Estonia, Finland, Greece, Hungary, Ireland, Italy, Lithuania, Latvia, Poland, Portugal, Romania, Slovenia, Spain, Sweden, France, Netherlands, Belgium, Luxembourg, Slovakia, Norway, Türkiye, Republic of North Macedonia, Iceland, Montenegro, Serbia