



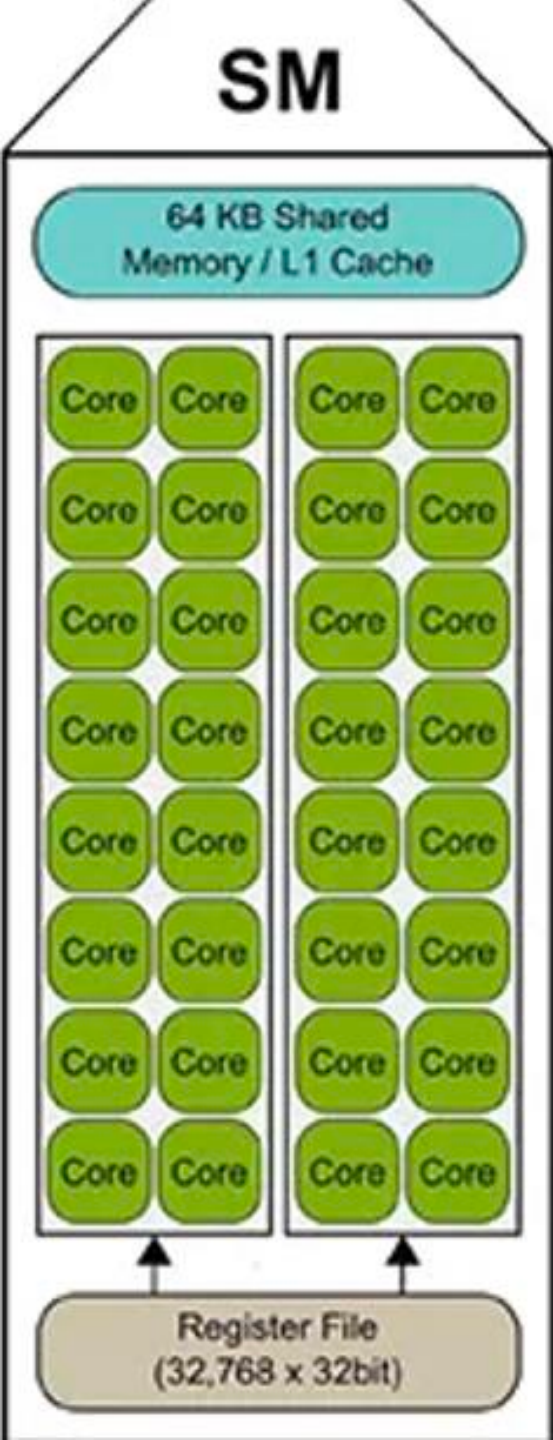
EURO²

Performance Engineering on CPUs and GPUs:

- GPUs: Things to be Careful for Performance -

Kamer Kaya, Sabancı University

Things to be Careful for Performance: (2) Using shared memory



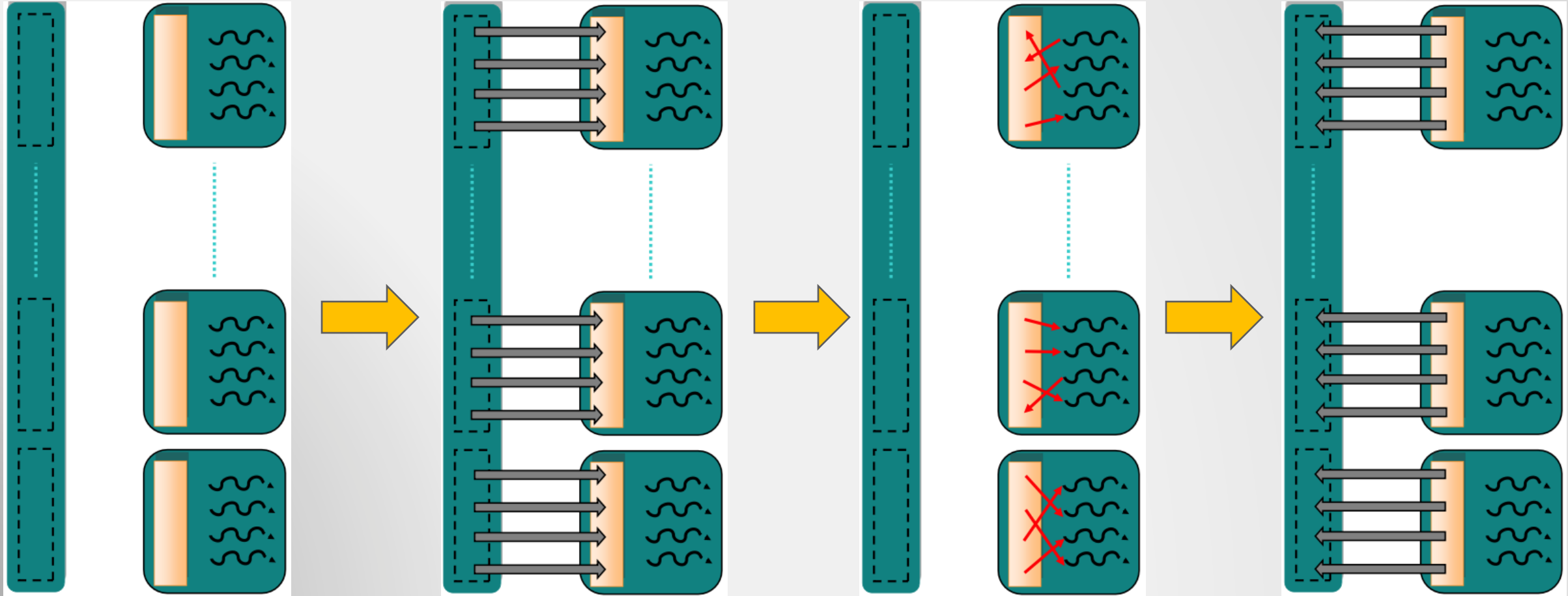
Shared memory is a fast, user-managed memory region located on the GPU chip.

It is accessible by all threads within the same thread block. It is much faster than global memory but limited in size. It has;

Low Latency: Significantly reduces memory access times compared to global memory.

Banked Access: Organized into memory banks to allow parallel access by threads (we will come to this later).

Things to be Careful for Performance: (2) Using shared memory



Things to be Careful for Performance:

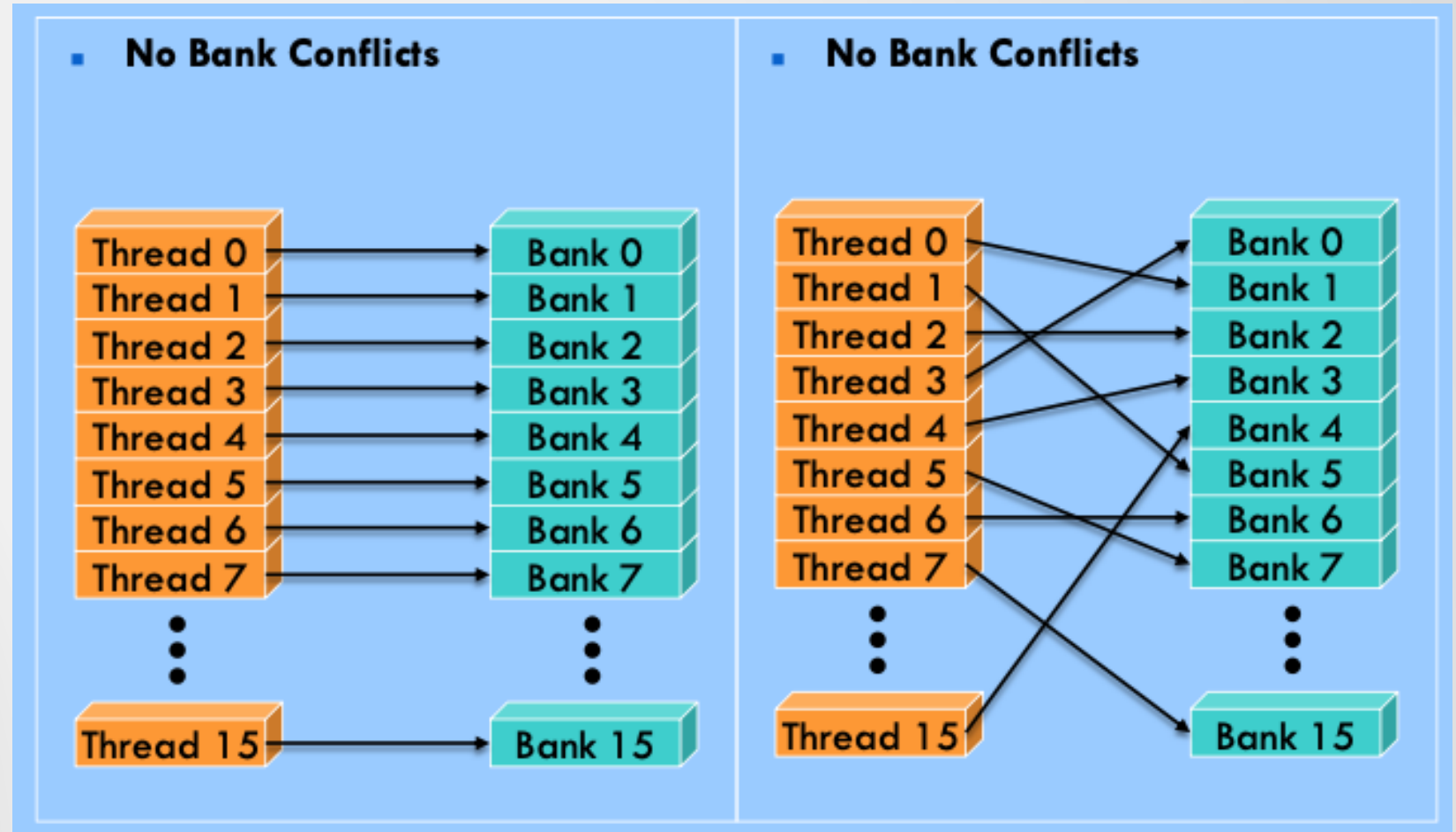
(3) Bank conflicts

- Shared memory is banked
 - Only matters for threads within a warp
 - Full performance with some restrictions
 - Threads can each access different banks
 - Or can all access the same value
- Consecutive words are in different banks
- If two or more threads access the same bank but different value, we get bank conflicts

Things to be Careful for Performance:

(3) Bank conflicts

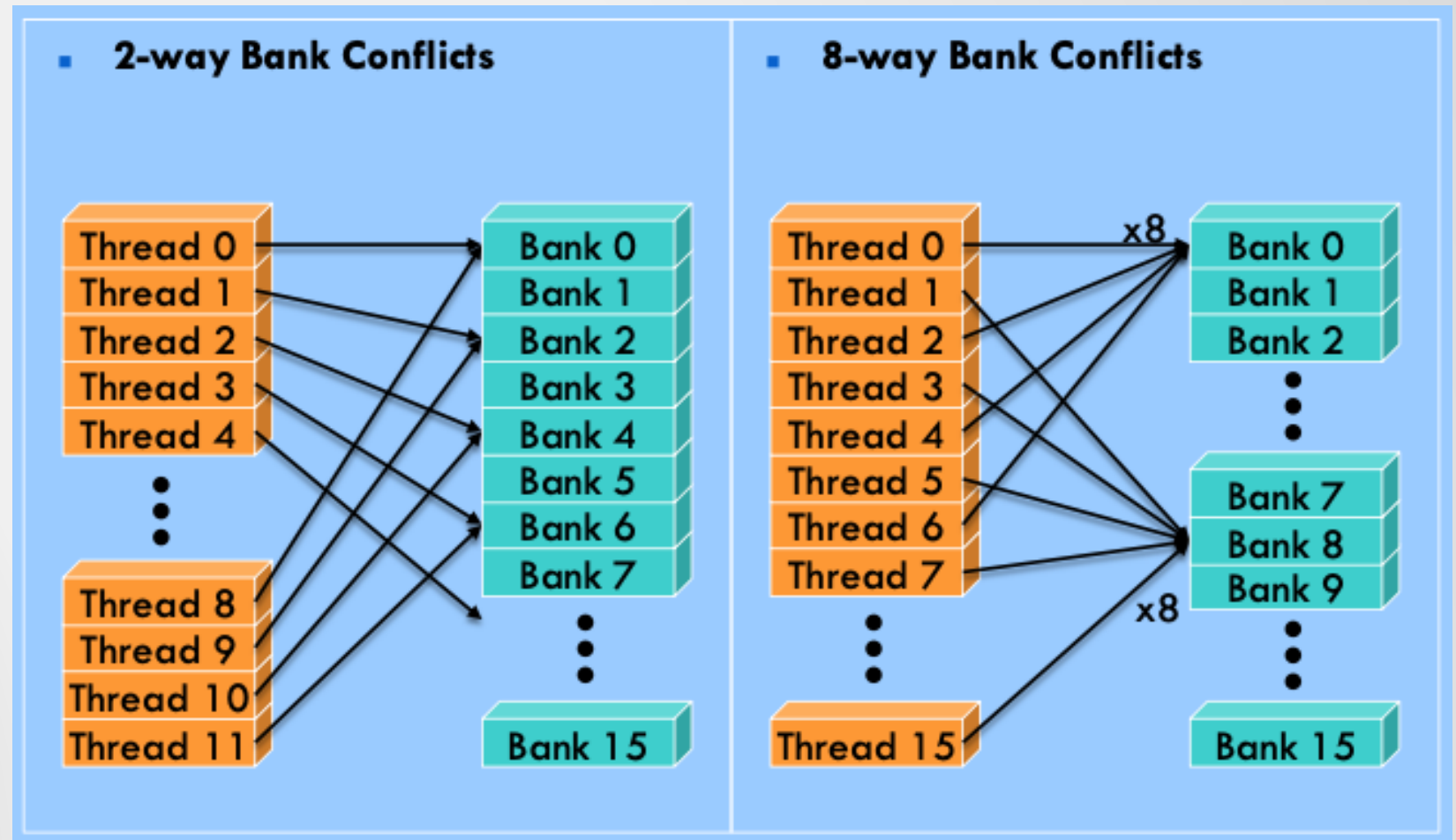
- When every thread accessed a different bank, there is no bank conflict.



Things to be Careful for Performance:

(3) Bank conflicts

- When different items are requested from the same bank, it is a **conflict**.



Things to be Careful for Performance:

(3) Bank conflicts

Let's check some examples on shared memory usage and bank conflicts.

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