Contents

10.3.2.2 Code Fusion

Code Fusion is an **optimization technique** used to improve the performance of programs, particularly in parallel computing. It involves **combining multiple operations into a single**, more efficient operation. By doing so, code fusion increases arithmetic intensity, reduces memory/cache usage, and enhances overall computational efficiency.

Key Points

- **Combining Operations**. Code fusion ***fuses* together multiple operations so that they are performed simultaneously**, rather than sequentially.

  Example 24: Code Fusion

  Instead of performing a series of separate operations on an array, we combine them into one loop, reducing the number of passes over the data.

- **Increasing Arithmetic Intensity**. Arithmetic intensity refers to the **ratio of computational operations to memory operations**. By fusing operations, we increase this ratio, meaning **more calculations are done per memory access**.

  Benefit. Higher arithmetic intensity leads to better utilization of the CPU and other computational resources.

- **Reducing Memory/Cache Usage**. By reducing the number of intermediate results that need to be stored in memory, code fusion **minimizes memory access and optimizes cache usage**.

- **Use of Registers**. Ideally, **operations can be performed using registers alone**, which are the fastest form of storage in a CPU. By keeping data in registers and reducing memory access, code fusion achieves maximum computational efficiency.

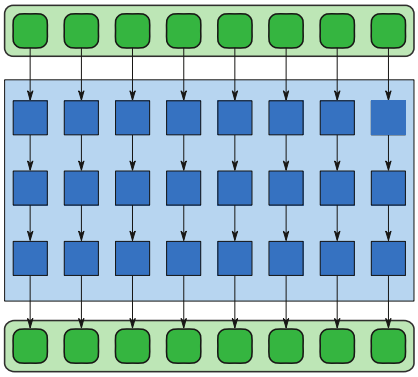

Figure 50: Graphical example of a code fusion.