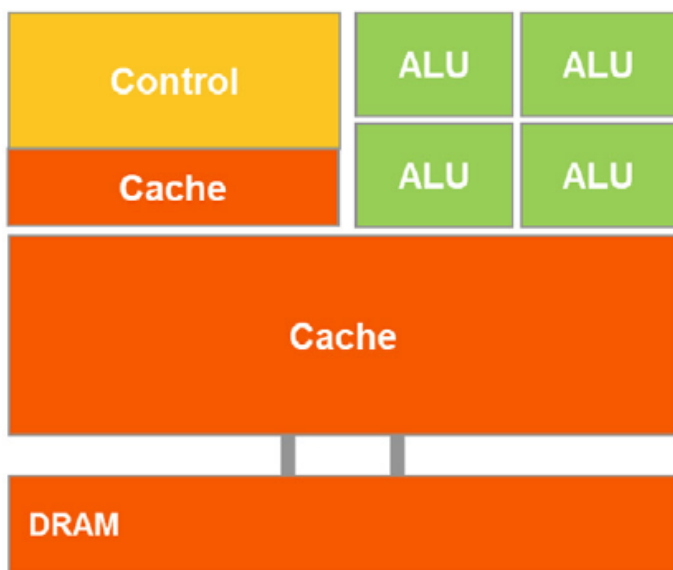


# CHAPTER 1

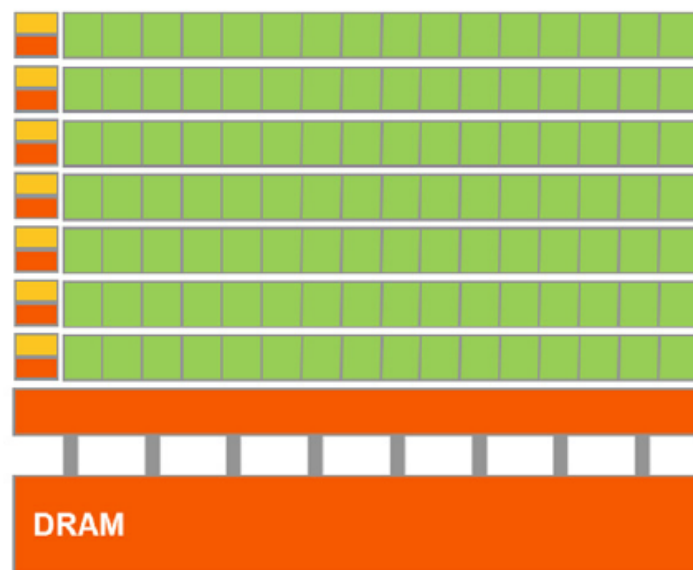
## Introduction



## CPU

(latency-oriented design)

(A)



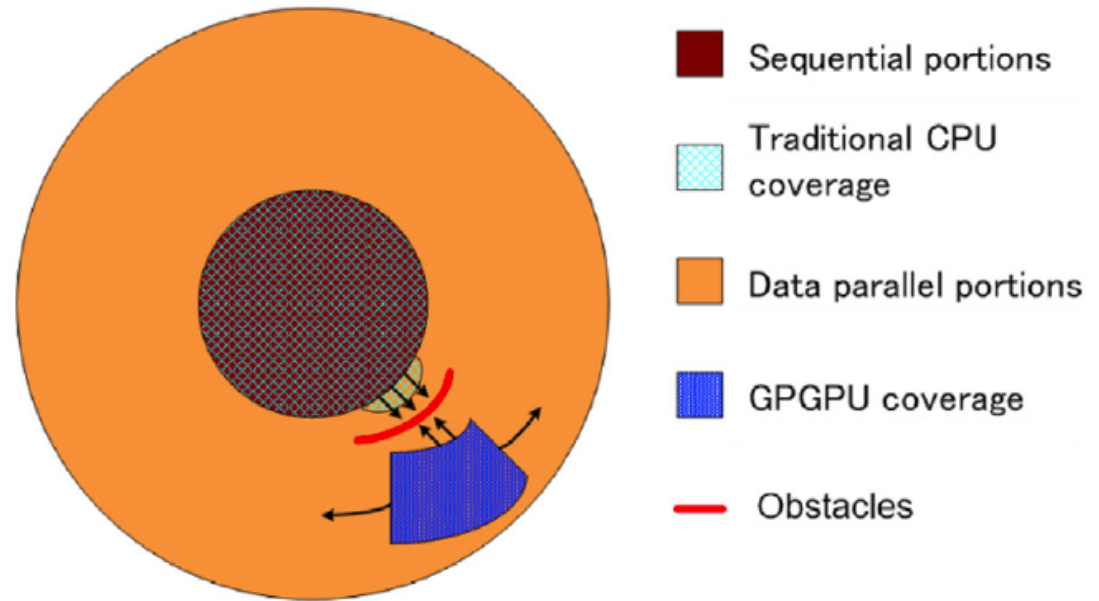
## GPU

(throughput-oriented design)

(B)

**FIGURE 1.1**

CPUs and GPUs have fundamentally different design philosophies: (A) CPU design is latency oriented; (B) GPU design is throughput-oriented.



**FIGURE 1.2**

Coverage of sequential and parallel application portions. The sequential portions and the traditional (single-core) CPU coverage portions overlap with each other. The previous GPGPU technique offers very limited coverage of the data parallel portions, since it is limited to computations that can be formulated into painting pixels. The obstacles refer to the power constraints that make it hard to extend single-core CPUs to cover more of the data parallel portions.