Operations in each Pipeline Stage

Clock Pulse Number	Segment 1		Segment 2			Segment 3
	R1	R2	R3	R4	R5	
1	A1	B1				
2	A2	B2		A1 * B1	C1	
3	A3	B3		A2 * B2	C2	A1 * B1 + C1
4	A4	B4		A3 * B3	C3	A2 * B2 + C2
5	A5	B5		A4 * B4	C4	A3 * B3 + C3
6	A6	B6		A5 * B5	C5	A4 * B4 + C4
7	A7	B7		A6 * B6	C6	A5 * B5 + C5
8				A7 * B7	C7	A6 * B6 + C6
9						A7 * B7 + C7

General Considerations

- Any operation that can be decomposed into a sequence of suboperations of about the same complexity can be implemented by a pipeline processor.

- The technique is efficient for those applications that need to repeat the same task many times with different sets of data.

- The general structure of a four-segment pipeline is illustrated in Fig. 9-3.

General Considerations

- The operands pass through all four segments in a fixed sequence.
- Each segment consists of a combinational circuit Si that performs a suboperation over the data stream flowing through the pipe.
- The segments are separated by registers Ri that hold the intermediate results between the stages.

General Considerations

- Information flows between adjacent stages under the control of a common clock applied to all the registers simultaneously.

- We task define a task as the total operation performed going through all the segments in the pipeline.

General Considerations

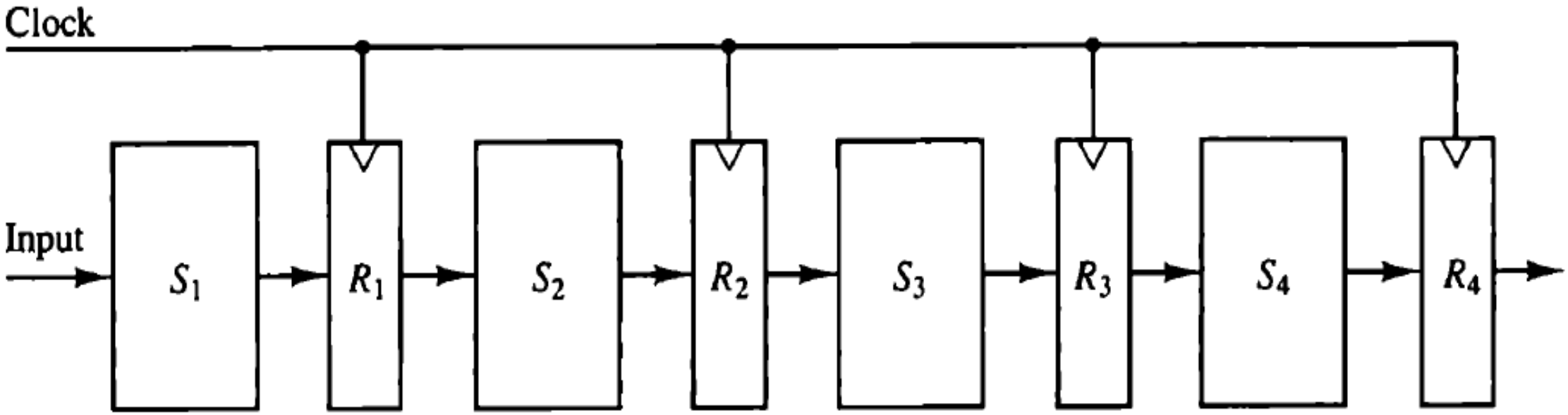


Figure 9-3 Four-segment pipeline.

space- time diagram

- The behavior of a pipeline can be illustrated with a space-time diagram.
- This is a diagram that shows the segment utilization as a function of time.
- The space-time diagram of a four-segment pipeline is demonstrated in Fig. 9-4.

space- time diagram

- The horizontal axis displays the time in clock cycles and the vertical axis gives the segment number.

- The diagram shows six tasks T1 through T6 executed in four segments.

- Initially, task 1i is handled by segment 1.

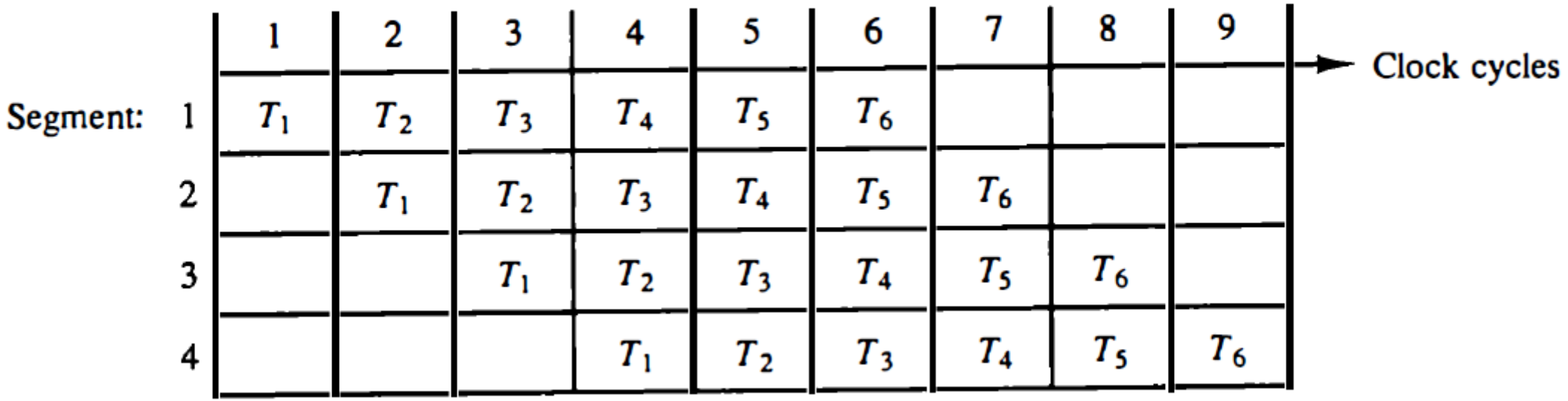
space- time diagram

- After the first clock, segment 2 is busy with T1, while segment 1 is busy with task T2.
- Continuing in this manner, the first task T1 is completed after the fourth clock cycle.
- From then on, the pipe completes a task every clock cycle.

space- time diagram

- No matter how many segments there are in the system, once the pipeline is full, it takes only one clock period to obtain an output.

Figure 9-4 Space-time diagram for pipeline.



space- time diagram

- Now consider the case where a k-segment pipeline with a clock cycle time tp is used to execute n tasks.

- The first task T1 requires a time equal to ktp to complete its operation since there are k segments in the pipe.

space- time diagram

- The remaining n - 1 tasks emerge from the pipe at the rate of one task per clock cycle and
- they will be completed after a time equal to (n - 1)tp.

- Therefore, to complete n tasks using a k-segment pipeline requires k + (n - 1) clock cycles.

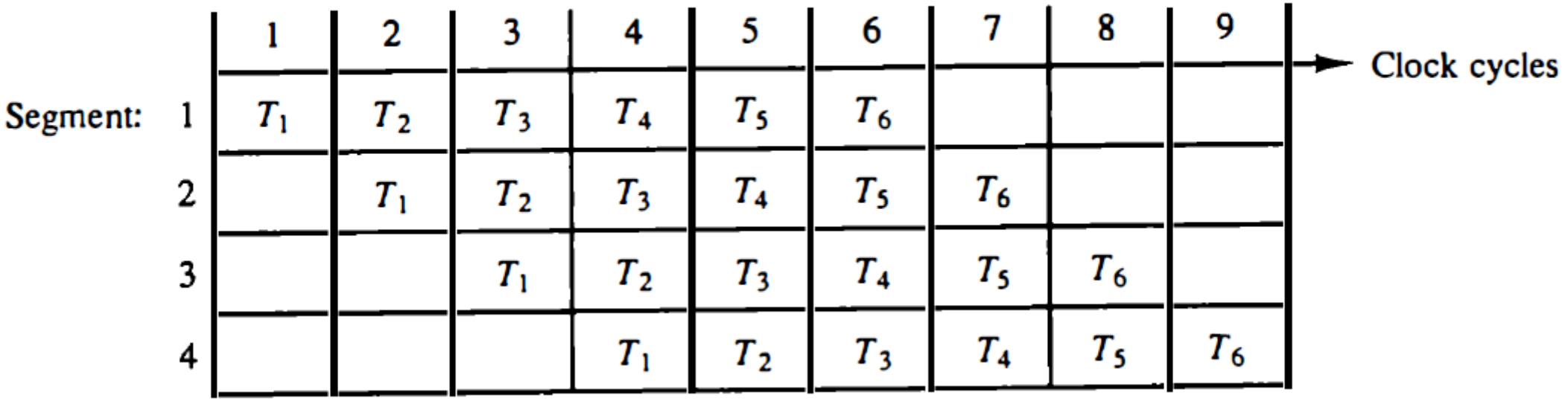
space- time diagram

- the diagram of Fig. 9-4 shows four segments and six tasks.

- The time required to complete all the operations is 4 + (6 - 1) = 9 clock cycles, as indicated in the diagram.

space- time diagram

Figure 9-4 Space-time diagram for pipeline.



space- time diagram

- Next consider a non pipeline unit that performs the same operation and takes a time equal to tn to complete each task.

- The total time required for n tasks is ntp.

- The speedup of a pipeline processing over an equivalent nonpipelined processing is defined by the ratio

$$S = \frac{nt_n}{(k + n - 1)t_p}$$

space- time diagram

As the number of tasks increases, n becomes much larger than $k - 1$, and $k + n - 1$ approaches the value of n. Under this condition, the speedup becomes

$$S = \frac{t_n}{t_p}$$

space- time diagram

If we assume that the time it takes to process a task is the same in the pipeline and nonpipeline circuits, we will have $t_n = kt_p$. Including this assumption, the speedup reduces to

$$S = \frac{kt_p}{t_p} = k$$

This shows that the theoretical maximum speedup that a pipeline can provide is k, where k is the number of segments in the pipeline.

space- time diagram

- To duplicate the theoretical speed advantage o f a pipeline process by means of multiple functional units, it is necessary to construct k identical units that will be operating in parallel.

- The implication is that a k-segment pipeline processor can be expected to equal the performance of k copies of an equivalent nonpipeline circuit under equal operating conditions.

space- time diagram

- This is illustrated in Fig. 9-5, where four identical circuits are connected in parallel.
- Each P circuit performs the same task of an equivalent pipeline circuit.
- Instead of operating with the input data in sequence as in a pipeline, the parallel circuits accept four input data items simultaneously and perform four tasks at the same time.

space- time diagram

- As far as the speed of operation is concerned, this is equivalent to a four segment pipeline.

- Note that the four-unit circuit of Fig. 9-5 constitutes a single-instruction multiple-data (SIMD) organization since the same instruction is used to operate on multiple data in parallel.

space- time diagram

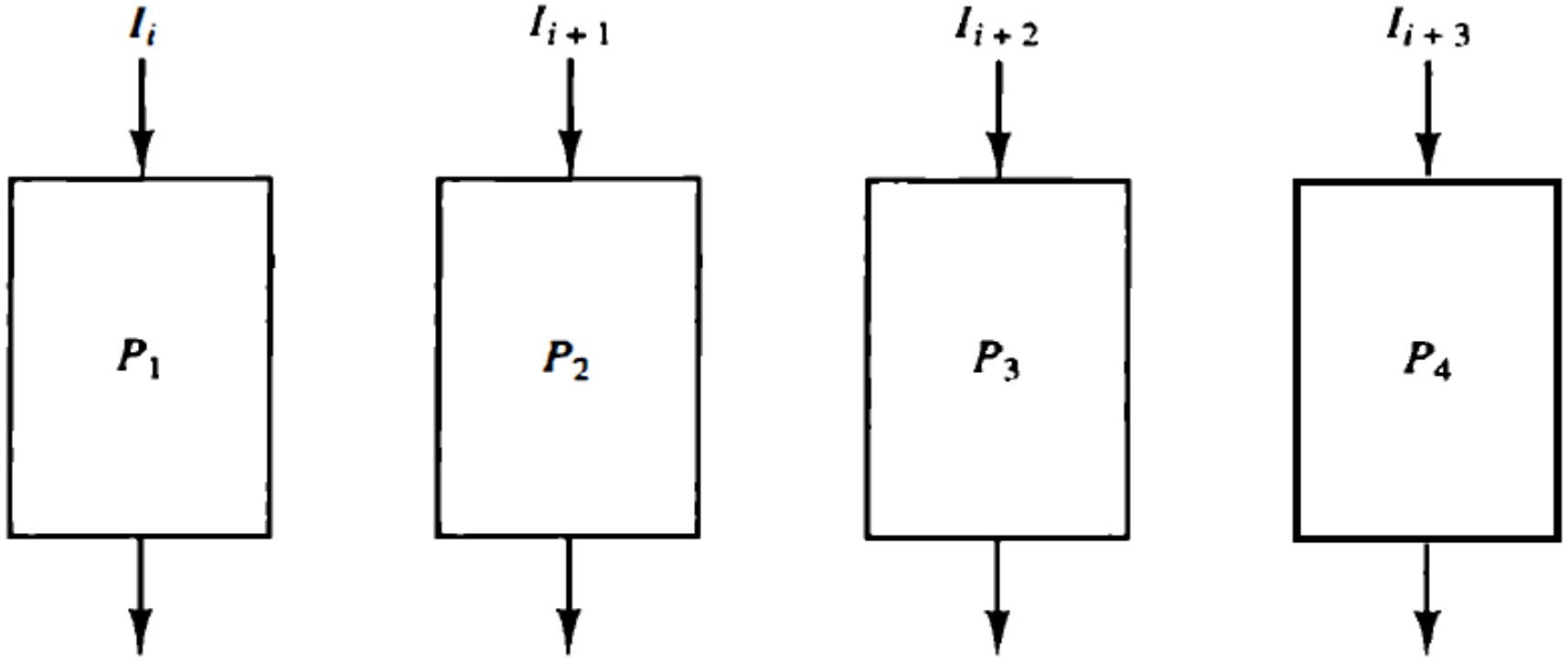


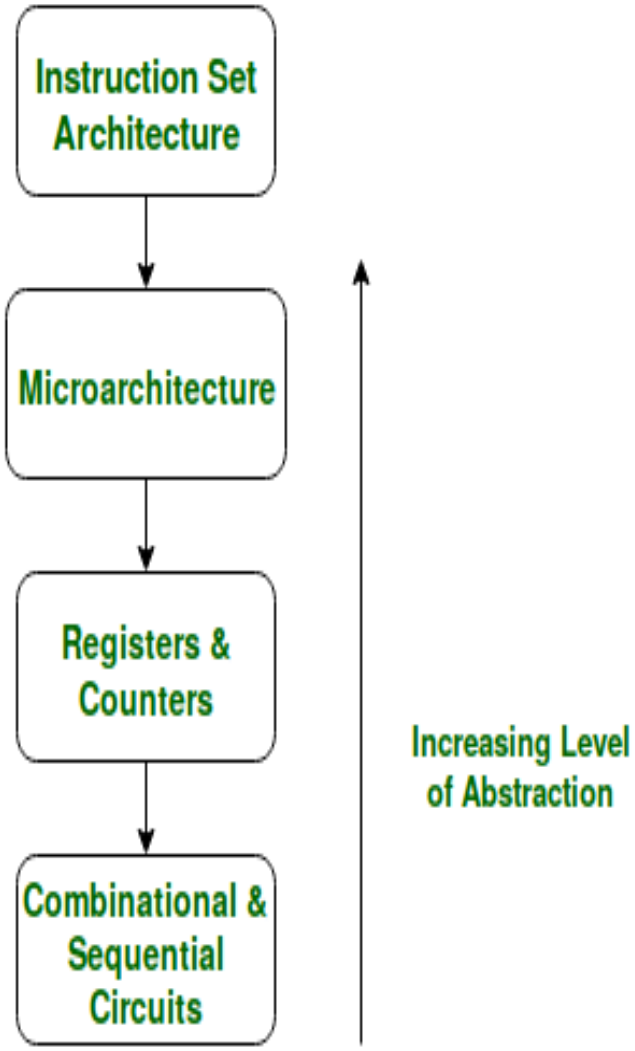
Figure 9-5 Multiple functional units in parallel.

Micro-architecture

- A microarchitecture is a hardware implementation of an ISA (instruction set architecture).
- An ISA is a structure of commands and operations used by software to communicate with hardware.
- A microarchitecture is the hardware circuitry that implements one particular ISA.

Micro-architecture

- For example, x86-64 is the ISA used by most modern laptop and desktop computers.
- It is implemented by various microarchitectures, including those designed by Intel.
- Software that is compiled for the x86-64 ISA can run on any microarchitecture designed to use the x86-64 instruction set.



Micro-architecture

- The ISA defines the types of instructions to be supported by the processor. Based on the type of operations they perform:-

- Arithmetic/Logic Instructions:

- These Instructions perform various Arithmetic & Logical operations on one or more operands.

Micro-architecture

- Data Transfer Instructions:

- These instructions are responsible for the transfer of instructions from memory to the processor registers and vice versa.

- Branch and Jump Instructions:

- These instructions are responsible for breaking the sequential flow of instructions and jumping to instructions at various other locations,

- this is necessary for the implementation of functions and conditional statements.

Micro-architecture

- The ISA defines the maximum length of each type of instruction, each instruction must be accommodated within 32 bits.
- The ISA defines the Instruction Format of each type of instruction.
- The Instruction Format determines how the entire instruction is encoded within 32 bits.
- Each Instruction Format has different instruction encoding schemes, and hence need to be interpreted differently by the processor.

Terminologies associated with Processors :

- Multithreading is a form of parallelization or dividing up work for simultaneous processing.

- Instead of giving a large workload to a single core, threaded programs split the work into multiple software threads.

- These threads are processed in parallel by different CPU cores to save time.

Terminologies associated with Processors :

- X86 architecture The x86 is developed based on the Intel 8086 microprocessor and its 8088 variant where it started out as a 16-bit instruction set for 16-bit processors,

- where many additions and extensions have been added to the x86 where it grew to 32-bit instruction sets over the years with almost entirely full backward compatibility.

Graphics Processing Unit

- Like a motherboard, a graphics card is a printed circuit board that houses a processor and RAM. It also has an input/output system (BIOS) chip, which stores the card's settings and performs diagnostics on the memory, input and output at startup.

- A graphics card's processor, called a graphics processing unit (GPU), is similar to a computer's CPU. A GPU, however, is designed specifically for performing the complex mathematical and geometric calculations that are necessary for graphics rendering. Some of the fastest GPUs have more transistors than the average CPU. A GPU produces a lot of heat, so it is usually located under a heat sink or a fan.

Graphics Processing Unit

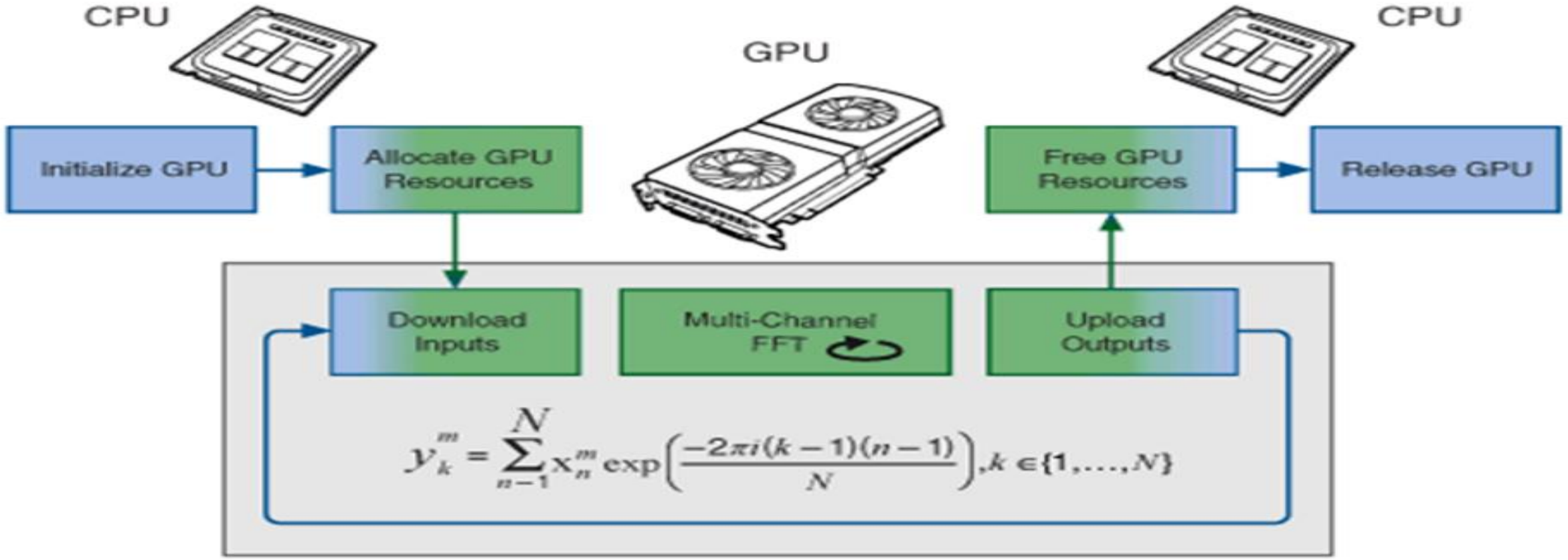
- A GPU however is more dedicated in function. It takes that same function that a CPU was processing and completes it all at once.

- The specified function requested of a GPU enters the GPU's hundred's of cores, and processes all at a single point in time, where it handles each process parallel to the next.

Graphics Processing Unit

- The GPU has become the most powerful processing unit of the system, especially since organizations are beginning to rely more on the processing power of a GPU rather than a CPU for the single fact that it can process more at a single point in time faster than a CPU.

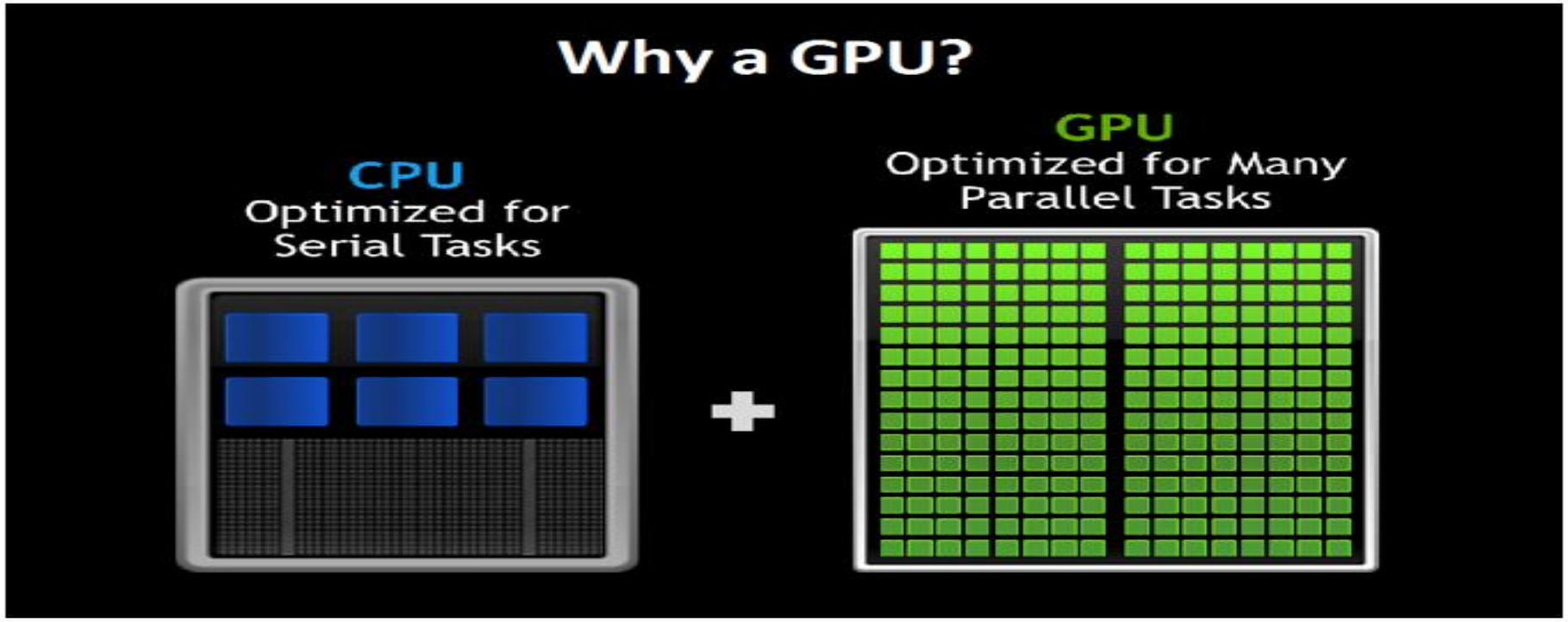
Graphics Processing Unit



Graphics Processing Unit

There are two different types of GPUs:

- **Integrated GPUs** are located on a PC's CPUand share memory with the CPU processor.
- **Discrete GPUs** live on their own card and have their own video memory (VRAM), so that the PC doesn't have to use its RAM for graphics.



ARM processors

- Advanced RISC Machine (ARM) Processor is considered to be family of Central Processing Units that is used in music players, smart phones, tablets and other consumer electronic devices.

- The architecture of ARM processor is created by Advanced RISC Machines, hence name ARM. This needs very few instruction sets and transistors.

ARM processors

- It has very small size.

- This is reason that it is perfect fit for small size devices. It has less power consumption along with reduced complexity in its circuits.

ARM processors

- They can be applied to various designs such as 32-bit devices and embedded systems.

- They can even be upgraded according to user needs.

ARM processors

- The main features of ARM Processor are mentioned below :

- Multiprocessing Systems –ARM processors are designed so that they can be used in cases of multiprocessing systems where more than one processors are used to process information.

- First processor introduced by name of ARMv6K had ability to support 4 CPUs along with its hardware.

ARM processors

- Tightly Coupled Memory –Memory of ARM processors is tightly coupled.

- This has very fast response time. It has low latency (quick response) that can also be used in cases of cache memory being unpredictable.

ARM processors

- Memory Management –ARM processor has management section.
- This includes Memory Management Unit and Memory Protection Unit.
- These management systems become very important in managing memory efficiently.

•**One cycle execution time –**ARM processor is optimised for each instruction on CPU. Each instruction is of fixed length that allows time for fetching future instructions before executing present instruction. ARM has CPI (Clock Per Instruction) of one cycle.

•**Pipelining–**Processing of instructions is done in parallel using pipelines. Instructions are broken down and decoded in one pipeline stage. The pipeline advances one step at a time to increase throughput (rate of processing).

ARM processors

- **Large number of registers** –Large number of registers are used in ARM processor to prevent large amount of memory interactions. Registers contain data and addresses. These act as local memory store for all operations.

- Advantages of ARM Processor :
- Affordable to create
- Low Power Consumption
- Work Faster
- Multiprocessing feature

ARM processors

- Disadvantages of ARM Processor :

- The speeds are limited in some processors which might create problems.

- There must be proper execution of instructions by programmer.

- This is because entire performance of ARM processors depend upon their execution.

Multi-core Processors

A **multicore processor** is a single integrated circuit (chip multiprocessor or CMP) that contains multiple core processing units, more commonly known as *cores*.

There are many different multicore processor architectures, which vary in terms of

Number of cores. Different multi core processors often have different numbers of cores. For example, a quad-core processor has four cores. The number of cores is usually a power of two.

ARM processors

- **Number of core types**.

  - **Homogeneous (symmetric) cores.** All of the cores in a homogeneous multicore processor are of the same type; typically the core processing units are general-purpose central processing units that run a single multicore operating system.

  - **Heterogeneous (asymmetric) cores**. Heterogeneous multicore processors have a mix of core types that often run different operating systems and include graphics processing units.

System on Chip (SoC)based architectures

- Mobile device processor architecture became simple with SOC designs.
- Real time responsiveness in mobile devices can be managed by using an enhanced DSP hybrid chip.
- Lowering the voltage of the chip enables low power operation in mobile devices.

System on Chip (SoC)based architectures

- Qualcomm Snapdragon Processors Snapdragon is a family of mobile system on a chip (SoC) processor architecture provided by Qualcomm.

- Scorpion, the original snapdragon CPU had many features similar to ARM Cortex-A8 core based on ARMv7 instruction set,

- but with an added advantage of higher performance utilizing SIMD operations.

Smartphone Hardware Architecture

- Every modern smartphone today uses a System on a Chip (SoC) Architecture with the following 3 primary components:

- - Application processor executing the user's application software with instructions from theoperatingsystem(OS).

- A baseband (or modem) processor with its own OS components performing baseband radio transmission and reception of audio, video and data

- - Various peripheral devices for the user interface.

Basic Smartphone architecture