Integrated circuit design based on CMOS technology principle and its application in GPU

Jiahao Chen

School of Electrical Engineering, Anhui Polytechnic University, Wuhu, Anhui, China, 241000

1641308688@qq.com

Abstract. In today's society, the application of integrated circuit technology can be seen everywhere, especially in the past two decades. This paper mainly studies the principle and design of CMOS devices in IC technology and discusses the research and analysis of the acceleration algorithm of IC design. This paper adopts the research method of literature review and analysis to summarize the existing research results. This paper first introduces the development background of integrated circuit technology and the importance of CMOS technology. Subsequently, the concept and interconnection principle of CMOS device, and the combined circuit design and sequential logic circuit design principle of dynamic and static CMOS are explained in detail. Then, the application principle of CMOS technology in GPU is analyzed, and its specific application in GPU acceleration algorithm is analyzed. Finally, the application of CMOS technology in integrated circuits and its application and acceleration effect are summarized.

Keywords: CMOS technology, combined circuit design, GPU.

1. Introduction

The purpose of this paper is to study the principle and design of Complementary Metal Oxide Semiconductor (CMOS) devices and explore the acceleration algorithm of integrated circuit design. CMOS is a commonly used complementary metal oxide semiconductor technology that achieves high performance, low power consumption and reliability integrated circuit design by exploiting the complementary properties of p-and n-type transistors. With the wide development of electronic product applications, a deep understanding and optimization of CMOS device design is crucial to improving the performance of electronic products. However, there are still some gaps in the principle and design of CMOS devices, which requires further exploration. Therefore, this paper aims to fill the knowledge gap in this research field and systematically elaborate the CMOS device principle and design to provide a more comprehensive and effective integrated circuit design method. Specifically, this paper focuses on the basic concepts and interconnection principles of CMOS devices, including the structure, characteristics and working principles. In addition, it also focuses on the dynamic and static CMOS combined circuit design methods and their advantages and disadvantages, as well as the sequential logic circuit design principles. In addition, the paper also focuses on the application of CMOS technology in Graphic Processing Unit (GPU) and shows its effect on GPU acceleration algorithm through comparison. This paper adopts the research method of literature review and analysis, comprehensively

summarizes the existing research results, and comprehensively and systematically expounds the principle and design of CMOS device and the acceleration algorithm of integrated circuit design. The research in this paper provides a deep understanding of the principle and design of CMOS devices that can help optimize the performance and power consumption of electronic products. At the same time, this paper provides a reference for future researchers and helps to further explore new ideas and methods of calculating and optimizing integrated circuits.

2. Principle and design of CMOS devices

2.1. Basic concepts and interconnect principle of CMOS devices
CMOS consists of two types of metal oxide semiconductor field-effect transistors (MOSFET): P-type (pMOS) and N-type (nMOS). These MOSFETs have three electrodes: source, drain, and gate. By applying a voltage to the gate, the current flow between the drain and source is controlled. The N-type MOSFET conducts when the gate voltage exceeds a threshold, allowing current flow from drain to source. The P-type MOSFET conducts when the gate voltage is below a threshold, enabling current flow from source to drain. CMOS devices utilize both pMOS and nMOS to achieve flexible current control, high-speed, and low-power operation. Interconnection technology plays a vital role in CMOS design by connecting MOSFETs through metal wires or polysilicon, ensuring reliable circuit functionality. The layout of metal wires significantly affects circuit performance, considering factors such as length, resistance, volume, and interconnection layer planning. CMOS inverters are common devices in CMOS technology. See Figure 1 for the structure diagram of the CMOS reverser.
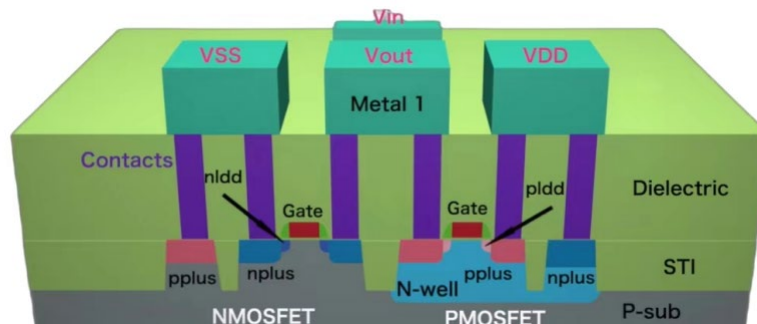

Figure 1. Structure diagram of the CMOS reverser.

2.2. Design of dynamic and static CMOS combined circuit
Dynamic and static CMOS combined circuit design is a technique that combines two design methods to achieve power-efficient and reliable circuit operation. In this approach, the main logic portion of the circuit is designed using dynamic CMOS, which enables low power consumption and high-performance logic operations. However, dynamic CMOS has limitations such as leakage current and timing issues.

To overcome these limitations, static CMOS is used in critical paths, storage components, and parts requiring high stability. Static CMOS has low power consumption when not switching and provides stable outputs. By combining dynamic and static CMOS, the circuit can maintain stable and reliable operation while achieving high performance and low power consumption.

This combined circuit design method is widely utilized in modern integrated circuits to meet the power consumption, performance, and reliability requirements in various application fields.

2.3. Time logic circuit design principle
By introducing the clock signal and the storage element, considering the change of the input signal and the clock trigger time, the logic function needs to operate according to a specific sequence. Design of a timing logic circuit requires comprehensively considering the clock period, edge trigger and state transition to ensure the correct timing operation and data stability. Timing logic circuit is often used in counter, state machine and controller applications, through reasonable design can achieve complex

timing and control function, and ensure the accuracy and stability of digital electronic system [1]. In Figure 2, a 3-bit Gray code synchronous addition counter composed of triggers is used.
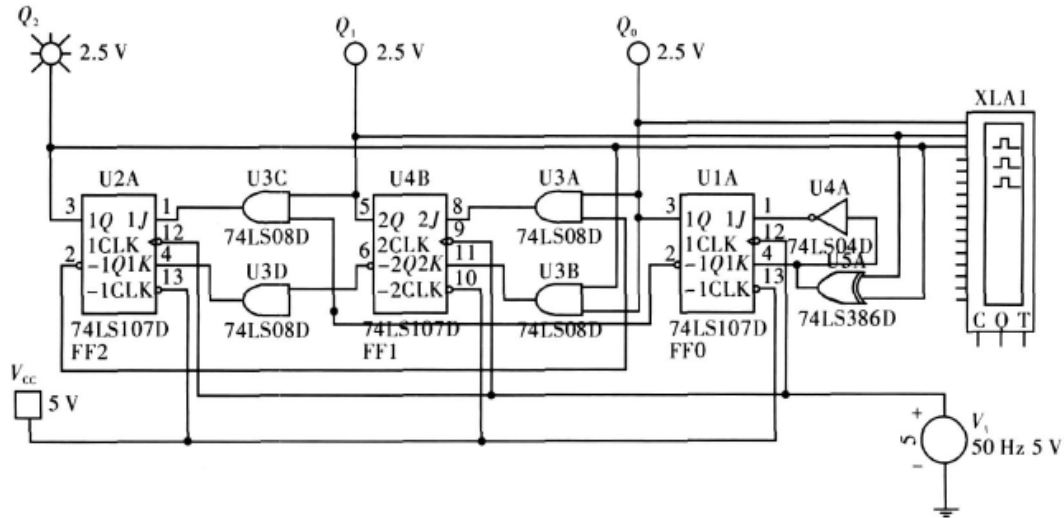

Figure 2. 3-bit Gray code synchronization addition counter [2].

2.4. Design principle and method of combined logic circuit

A digital logic circuit consists of combined logic and sequential logic. The design process involves two stages: translating logical functions from a literal description to a logical representation, and converting between different logical descriptions. Tools such as truth tables, logic expressions, circuit diagrams, and hardware description languages aid in this process. Design methods vary depending on the tools used. Traditional digital system design uses small and medium-sized logic devices to implement circuitry and achieve system functionality, though it can be complex, it aids in understanding unit circuit logic. PLD design defines logic, input/output terminals within the chip, reducing workload and difficulty in circuit diagram and board design. This enhances design flexibility and efficiency, making it an ideal device for modern digital systems [2].

3. Research and analysis of the acceleration algorithm for integrated circuit design

3.1. Principles of CMOS technology in GPU

The application principle of CMOS in GPU is mainly related to the manufacturing and operation of transistors and circuits. CMOS technology uses N and P-type transistors to construct the logic gates and storage units in the GPU. Among them, N-type transistor is a device with good conductivity, while P-type transistor has poor conductivity. Together, the two transistors form a complementary structure that allows the circuit in the GPU to operate on both positive and negative logic. The advantage of this structure is that power is consumed only when performing logical operations, with no current flow being static. This makes CMOS technology very suitable for high-performance and low-power applications. The operation of the CMOS transistor controls the flow of the current by applying different voltages. This combination of positive and negative logic enables the GPU to perform complex mathematical and logical operations. The large number of transistors in the GPU and the complex circuit connections between them make parallel processing possible. The GPU can handle multiple data and tasks simultaneously, dispersing the computing load over multiple processing units, thus improving processing speed and performance. Overall, CMOS technology endows GPU with powerful graphics processing power and high-performance computing power.

Most of the modern 3D graphics processing units (GPUs) are implemented as programmable parallel processors with extremely high computing power. The high computing capability of GPUs comes at the

cost of high power dissipation [3]. In the power gating technology, the CMOS technology controls the logic unit or functional module by controlling the on-off state of the transistors. CMOS technology provides the advantages of high integration, low power consumption and reliability, allowing power gating technology to flexibly control the power supply of chips or modules, thus reducing power consumption and improving energy efficiency. The conduction or cut-off state is controlled by adjusting its gate voltage. Power gating technology takes advantage of this feature to cut off power to prevent leakage current leakage when needed and reconnect power to restore function when needed. This power control can accurately control different parts of the power supply, thus achieving higher energy efficiency and optimized power consumption. Therefore, power gating technology is an important application based on CMOS technology. Figure 3 is a power supply gating circuit controlled by the head transistor.
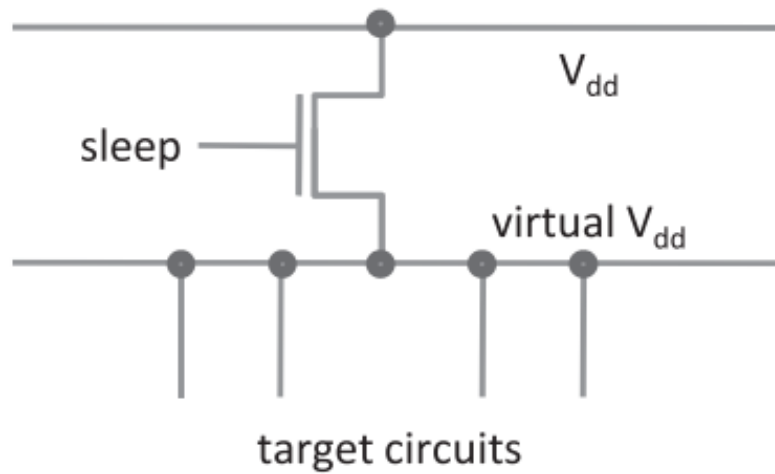

Figure 3. Power gating circuits controlled by header transistor [3].

3.2. Example analysis: the application of GPU acceleration algorithm
CUDA is a parallel computing platform and application programming interface (API) developed by NVIDIA. CUDA can use the GPU to accelerate compute-intensive tasks, such as deep learning, image processing, and so on. The hardware architecture of a modern GPU can be seen as a series of single-instruction multi-data (SIMD) multiprocessors, each handling a set of instructions in different memory components during a clock cycle [4].

JACKET Engine Is a MATLAB-based GPU computing engine developed by AccelerEyes. See Figure 4, at present, JACKET is only CUDA technology based on NVIDIA, and the computing engine of JACKET integrated graphical toolkit can fully play all the Open-GL functions of GPU. This integration of computing and visualization enables the development of real visual computing, and JACKET allows the M language and MATLAB to transparently adapt to GPU computing [5].
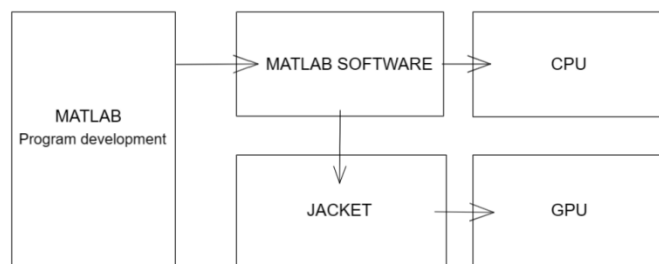

Figure 4. The MATLAB program development process based on JACKET [6].

At present, the mainstream CPU circulating in the market generally contains 2~8 cores, among which each core contains 3~6 execution lines. Most of these CPUs use improved instruction-level parallel technology and large-capacity caching to improve performance. Since most of the area on the chip is used to integrate the cache and control logic, there is a limit to the number of integrated cores. The current GPU produced by NVIDIA company generally contains 1~30 SM. The large number of execution units still makes the GPU floating point processing capacity much stronger than the CPU. As shown in Figure 5, we can see the difference in the computational power of the two.
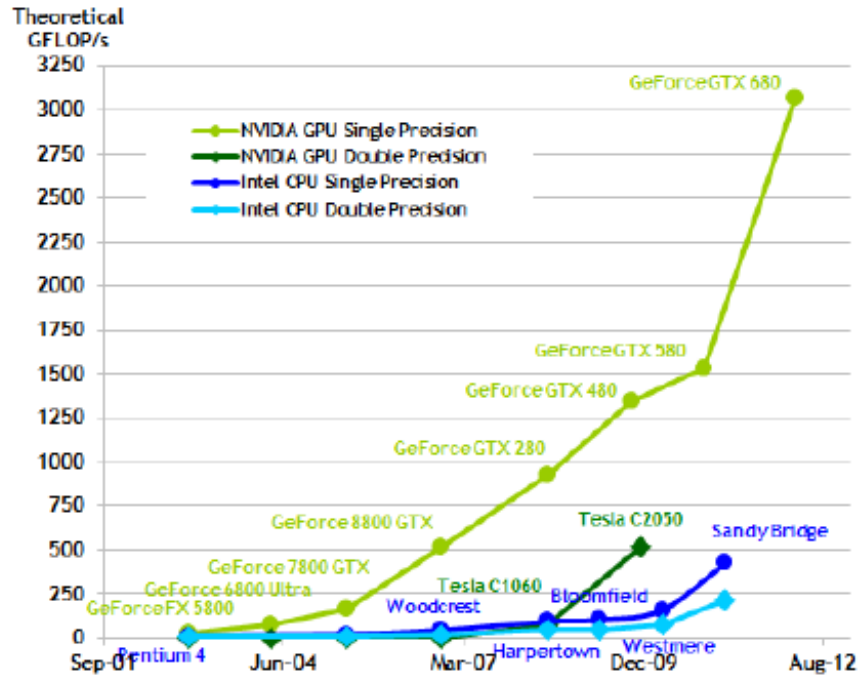

Figure 5. Comparison of peak floating point computational power between GPU and CPU [7].

4. Conclusion

This paper reviews the development of IC technology from the earliest discrete elements to the rapid development of modern IC. The importance of CMOS (complementary metal-oxide-semiconductor) technology in integrated circuit design, including low power consumption, high integration and good anti-interference ability, are highlighted. The basic principles of CMOS devices are described, covering key concepts such as structure, working principle and interconnection techniques. Other aspects related to CMOS devices such as material selection, process techniques, and device layout are also discussed. The combined circuit design method of dynamic and static CMOS are thoroughly studied. Meanwhile, the advantages and disadvantages of these two design methods are also compared and evaluated. This paper clearly obtains the successful application of this technology, and compares the algorithm speed to some extent, and obtains the advantages of the technology product GPU. In this paper, there is no detailed discussion on the application and speed of relevant specific algorithms, and the difference in GPU call speed between different algorithms. In the future, integrated circuit technology will develop towards smaller, higher performance, lower power consumption and higher integration, and provide innovation and solutions for various industries.

References

[1] CAI Xiancheng. Design and simulation of the timing logic circuit based on Multisim [J]. Electronic Technology, 2010,23(12):12-13+17. DOI: 10.16180/j.cnki.issn1007-7820.201 0.12.009.

[2]   Zhang Jie, Sun Bin, Cui Ruixue, etc. Design method and design skills of combinatorial logic circuit [J]. Journal of North China Institute of Aerospace Technology, 2014,24 (04): 23-26.

[3]   Po-Han Wang, Chia-Lin Yang, Yen-Ming Chen, and Yu-Jung Cheng. 2011. Power gating strategies on GPUs. ACM Trans. Archit. Code Optim. 8, 3, Article 13 (October 2011), 25 pages. https://doi.org/10.1145/2019608.2019612

[4]   Nvidia.CUDA4,1, [M]. California: Nvidia, 2011:5-8.

[5]   AccelerEyes. JACKET1.7 GettingStartedGuide [M. Atlanta: AccelevEyes, 2011: 8-12.

[6]   Wang Heng, building from a hillhouse. A MATLAB calculation and simulation study based on the GPU [J]. Journal of Guizhou University (Natural Science Edition), 2012,29(06):95-98+112. DOI: 10.15958/j.cnki.gdxbzrb. 2012.06.013.

[7]   Chen Qian. Research on the parallel algorithm for digital image processing based on GPU [D]. Graduate School of Chinese Academy of Sciences (Xi'an Institute of Optics and Fine Mechanics), 2014.