国外电子信息精品著作（影印版）

视觉感知的模拟超大规模集成电路实现

Analog VLSI Circuits —— Perception of Visual Motion

Stocker A. A.

科学出版社
www.sciencep.com

国外电子信息精品著作（影印版）

视觉感知的模拟超大规模集成电路实现

Stocker A. A.

科学出版社
北京

图字：01-2006-7267

内 容 简 介

计算神经系统科学是一个正在兴起的研究领域，近年来已经成为许多国家政府资助的研究方向，吸引着许多青年研究人员。本书分析了视觉运动感知的计算问题、模拟网络的优化方法等，最有特色处在于从大规模集成电路实现的角度分析了视觉运动处理的原理和算法，可以借助大规模集成电路的高集成度、低成本等优势，进行模拟的并行视觉运动感知。

本书的专业性很强，所涉及的问题非常前沿，属于交叉学科，极具发展潜力，其权威性不言而喻。对于从事神经网络、人工智能、控制理论的等领域的研究者，本书有很大的参考价值。在我国，这方面的研究还处于起步阶段，本书提出的视觉运动集成电路实现方法无疑是一种崭新的思路。

图书在版编目(CIP)数据

视觉感知的模拟超大规模集成电路实现：英文/（美）斯多克编著. —影印版. —北京：科学出版社，2007.1
（国外电子信息精品著作）

ISBN 978-7-03-018254-8

Ⅰ. 视… Ⅱ. 斯… Ⅲ. 人工神经元网络-超大规模集成电路-英文 Ⅳ. TN47

中国版本图书馆 CIP 数据核字(2006)第 149504 号

责任编辑：余 丁/责任印制：安春生/封面设计：陈 敬

科学出版社 出版
北京东黄城根北街 16 号
邮政编码：100717
http://www.sciencep.com

天时彩色印刷有限公司印刷

科学出版社发行 各地新华书店经销

*

2007 年 1 月第 一 版 开本：B5(720×1000)
2007 年 1 月第一次印刷 印张：15
印数：1—3 000 字数：372 000

定价：38.00 元

（如有印装质量问题，我社负责调换〈双青〉）

《国外电子信息精品著作》序

20 世纪 90 年代以来，信息科学技术成为世界经济的中坚力量。随着经济全球化的进一步发展，以微电子、计算机、通信和网络技术为代表的信息技术，成为人类社会进步过程中发展最快、渗透性最强、应用面最广的关键技术。信息技术的发展带动了微电子、计算机、通信、网络、超导等产业的发展，促进了生命科学、新材料、能源、航空航天等高新技术产业的成长。信息产业的发展水平不仅是社会物质生产、文化进步的基本要素和必备条件，也是衡量一个国家的综合国力、国际竞争力和发展水平的重要标志。在中国，信息产业在国民经济发展中占有举足轻重的地位，成为国民经济重要支柱产业。然而，中国的信息科学支持技术发展的力度不够，信息技术还处于比较落后的水平，因此，快速发展信息科学技术成为我国迫在眉睫的大事。

要使我国的信息技术更好地发展起来，需要科学工作者和工程技术人员付出艰辛的努力。此外，我们要从客观上为科学工作者和工程技术人员创造更有利于发展的环境，加强对信息技术的支持与投资力度，其中也包括与信息技术相关的图书出版工作。

从出版的角度考虑，除了较好较快地出版具有自主知识产权的成果外，引进国外的优秀出版物是大有裨益的。洋为中用，将国外的优秀著作引进到国内，促进最新的科技成就迅速转化为我们自己的智力成果，无疑是值得高度重视的。科学出版社引进一批国外知名出版社的优秀著作，使我国从事信息技术的广大科学工作者和工程技术人员能以较低的价格购买，对于推动我国信息技术领域的科研与教学是十分有益的事。

此次科学出版社在广泛征求专家意见的基础上，经过反复论证、仔细遴选，共引进了接近 30 本外版书，大体上可以分为两类，第一类是基础理论著作，第二类是工程应用方面的著作。所有的著作都涉及信息领域的最新成果，大多数是 2005 年后出版的，力求“层次高、内

容新、参考性强”。在内容和形式上都体现了科学出版社一贯奉行的严谨作风。

当然，这批书只能涵盖信息科学技术的一部分，所以这项工作还应该继续下去。对于一些读者面较广、观点新颖、国内缺乏的好书还应该翻译成中文出版，这有利于知识更好更快地传播。同时，我也希望广大读者提出好的建议，以改进和完善丛书的出版工作。

总之，我对科学出版社引进外版书这一举措表示热烈的支持，并盼望这一工作取得更大的成绩。

王越

中国科学院院士

中国工程院院士

2006 年 12 月

Foreword

Although we are now able to integrate many millions of transistors on a single chip, our ideas of how to use these transistors have changed very little from the time when John von Neumann first proposed the global memory access, single processor architecture for the programmable serial digital computer. That concept has dominated the last half century, and its success has been propelled by the exponential improvement of hardware fabrication methods reflected in Moore's Law. However, this progress is now reaching a barrier in which the cost and technical problems of constructing CMOS circuits at ever smaller feature sizes is becoming prohibitive. In future, instead of taking gains from transistor count, the hardware industry will explore how to use the existing counts more effectively by the interaction of multiple general and specialist processors. In this way, the computer industry is likely to move toward understanding and implementing more brain-like architectures.

Carver Mead, of Caltech, was one of the pioneers who recognized the inevitability of this trend. In the 1980s he and his collaborators began to explore how integrated hybrid analog–digital CMOS circuits could be used to emulate brain-style processing. It has been a hard journey. Analog computing is difficult because the physics of the material used to construct the machine plays an important role in the solution of the problem. For example, it is difficult to control the physical properties of sub-micron-sized devices such that their analog characteristics are well matched. Another problem is that unlike the bistable digital circuits, analog circuits have no inherent reference against which signal errors can be restored. So, at first sight, it appears that digital machines will always have an advantage over analog ones when high precision and signal reliability are required.

But why are precision and reliability required? It is indeed surprising that the industry insists on developing technologies for precise and reliable computation, despite the fact that brains, which are much more effective than present computers in dealing with real-world tasks, have a data precision of only a few bits and noisy communications.

One factor underlying the success of brains lies in their use of constraint satisfaction. For example, it is likely that the fundamental Gestalt Laws of visual perceptual grouping observed in humans arise from mechanisms that resolve and combine the aspects of an image that cohere from those that do not. These mechanisms rapidly bootstrap globally coherent solutions by quickly satisfying local consistency conditions. Consistency depends on relative computations such as comparison, interpolation, and error feedback, rather than absolute precision. And, this style of computation is suitable for implementation in densely parallel hybrid CMOS circuits.

The relevance of this book is that it describes the theory and practical implementation of constraint satisfaction networks for motion perception. It also presents a principled development of a series of analog VLSI chips that go some way toward the solution of some difficult problems of visual perception, such as the Aperture Problem, and Motion Segmentation.

These classical problems have usually been approached by algorithms, and simulation, suitable for implementation only on powerful digital computers. Alan Stocker's approach has been to find solutions suitable for implementation on a single or very small number of electronic chips that are composed predominantly of analog circuitry, and that process their visual input in real time. His solutions are elegant, and practically useful. The aVLSI design, fabrication, and subsequent analysis have been performed to the highest standards. Stocker discusses each of these phases in some detail, so that the reader is able to gain considerable practical benefit from the author's experience.

Stocker also makes a number of original contributions in this book. The first is his extension of the classical Horn and Schunck algorithm for estimation of two-dimensional optical flow. This algorithm makes use of a brightness and a smoothness constraint. He has extended the algorithm to include a 'bias constraint' that represents the expected motion in case the visual input signal is unreliable or absent. The second is the implementation of this algorithm in a fully functional aVLSI chip. And the third is the implementation of a chip that is able to perform piece-wise smooth optical flow estimation, and so is able (for example) to segment two adjacent pattern fields that have a motion discontinuity at their common boundary. The optical flow field remains smooth within each of the segmented regions.

This book presents a cohesive argument on the use of constraint satisfaction methods for approximate solution of computationally hard problems. The argument begins with a useful and informed analysis of the literature, and ends with the fine example of a hybrid motion-selection chip. This book will be useful to those who have a serious interest in novel styles of computation, and the special purpose hardware that could support them.

Rodney J. Douglas Zürich, Switzerland

Preface

It was 1986 when John Tanner and Carver Mead published an article describing one of the first analog VLSI visual motion sensors. The chip proposed a novel way of solving a computational problem by a collective parallel effort amongst identical units in a homogeneous network. Each unit contributed to the solution according to its own interests and the final outcome of the system was a collective, overall optimal, solution. When I read the article for the first time ten years later, this concept did not lose any of its appeal. I was immediately intrigued by the novel approach and was fascinated enough to spend the next few years trying to understand and improve this way of computation - despite being told that the original circuit never really worked, and in general, this form of computation was not suited for aVLSI implementations.

Luckily, those people were wrong. Working on this concept of collective computation did not only lead to extensions of the original circuit that actually work robustly under real-world conditions, it also provided me with the intuition and motivation to address fundamental questions in understanding biological neural computation. Constraint satisfaction provides a clear way of solving a computational problem with a complex dynamical network. It provides a motivation for the behavior of such systems by defining the optimal solution and dynamics for a given task. This is of fundamental importance for the understanding of complex systems such as the brain. Addressing the question *what* the system is doing is often not sufficient because of its complexity. Rather, we must also address the functional motivation of the system: *why* is the system doing what it does?

Now, another ten years later, this book summarizes some of my personal development in understanding physical computation in networks, either electronic or neural. This book is intended for physicists, engineers and computational biologists who have a keen interest in the computational question in physical systems. And if this book finally inspires a young graduate student to try to understand complex computational systems and the building of computationally efficient devices then I am very content – even if it takes another ten years for this to happen.

Acknowledgments

I am grateful to many people and institutions that have allowed me to pursue my work with such persistence and great scientific freedom. Foremost I want to thank my former advisor, Rodney Douglas, who provided me with a fantastic scientific environment in which many of the ideas originated that are now captured in this book. I am grateful for his encouragement and support during the writing of the book. Most of the circuits developments were performed when I was with the Institute of Neuroinformatics, Zürich Switzerland. My thanks go to all members of the institute at that time, and in particular to the late Jörg Kramer who introduced me to analog circuits design. I also want to thank the Swiss government, the Körber foundation, and the Howard Hughes Medical Institute for their support during the development and writing of this book.

Many colleagues and collaborators had a direct influence on the final form of this book by either working with me on topics addressed in this book or by providing invaluable suggestions and comments on the manuscript. I am very thankful to know and interact with such excellent and critical minds. These are, in alphabetical order: Vlatko Becanovic, Tobias Delbrück, Rodney Douglas, Ralph Etienne-Cummings, Jakob Heinzle, Patrik Hoyer, Giacomo Indiveri, Jörg Kramer, Nicole Rust, Bertram Shi, and Eero Simoncelli.

Writing a book is a hard optimization problem. There are a large number of constraints that have to be satisfied optimally, many of which are not directly related to work or the book itself. And many of these constraints are contradicting. I am very grateful to my friends and my family who always supported me and helped to solve this optimization problem to the greatest possible satisfaction.

Website to the book

There is a dedicated on-line website accompanying this book where the reader will find supplementary material, such as additional illustrations, video-clips showing the real-time output of the different visual motion sensor and so forth. The address is http://wiley.com/go/analog

The website will also contain updated links to related research projects, conferences and other on-line resources.

Contents

1

Introduction

Our world is a visual world. Visual perception is by far the most important sensory process by which we gather and extract information from our environment. Light reflected from objects in our world is a very rich source of information. Its short wavelength and high transmission speed allow us a spatially accurate and fast localization of reflecting surfaces. The spectral variations in wavelength and intensity in the reflected light resemble the physical properties of object surfaces, and provide means to recognize them. The sources that light our world are usually inhomogeneous. The sun, our natural light source, for example, is in good approximation a point source. Inhomogeneous light sources cause shadows and reflectances that are highly correlated with the shape of objects. Thus, knowledge of the spatial position and extent of the light source enables further extraction of information about our environment.

Our world is also a world of motion. We and most other animals are moving creatures. We navigate successfully through a dynamic environment, and we use predominantly visual information to do so. A sense of motion is crucial for the perception of our own motion in relation to other moving and static objects in the environment. We must predict accurately the relative dynamics of objects in the environment in order to plan appropriate actions. Take for example the following situation that illustrates the nature of such a perceptual task: the goal-keeper of a football team is facing a direct free-kick toward his goal.[1] In order to prevent the opposing team from scoring, he needs an accurate estimate of the real motion trajectory of the ball such that he can precisely plan and orchestrate his body movements to catch or deflect the ball appropriately. There is little more than just visual information available to him in order to solve the task. And once he is in motion the situation becomes much more complicated because visual motion information now represents the relative motion between himself and the ball while the important coordinate frame remains

[1]There are two remarks to make. First, "football" is referred to as the European-style football, also called "soccer" elsewhere. Second, there is no gender-specific implication here; a male goal-keeper was simply chosen so-as to represent the sheer majority of goal-keepers on earth. In fact, I particularly would like to include non-human, artificial goal-keepers as in robotic football (RoboCup [Kitano et al. 1997]).

Analog VLSI Circuits for the Perception of Visual Motion A. A. Stocker

static (the goal). Yet, despite its difficulty, with appropriate training some of us become astonishingly good at performing this task.

High performance is important because we live in a highly competitive world. The survival of the fittest applies to us as to any other living organism, and although the fields of competition might have slightly shifted and diverted during recent evolutionary history, we had better catch that free-kick if we want to win the game! This competitive pressure not only promotes a visual motion perception system that can determine quickly what is moving where, in which direction, and at what speed; but it also forces this system to be efficient. Efficiency is crucial in biological systems. It encourages solutions that consume the smallest amount of resources of time, substrate, and energy. The requirement for efficiency is advantageous because it drives the system to be quicker, to go further, to last longer, and to have more resources left to solve and perform other tasks at the same time. Our goal-keeper does not have much time to compute the trajectory of the ball. Often only a split second determines a win or a defeat. At the same time he must control his body movements, watch his team-mates, and possibly shout instructions to the defenders. Thus, being the complex sensory-motor system he is, he cannot dedicate all of the resources available to solve a single task.

Compared to human perceptual abilities, nature provides us with even more astonishing examples of efficient visual motion perception. Consider the various flying insects that navigate by visual perception. They weigh only fractions of grams, yet they are able to navigate successfully at high speeds through a complicated environments in which they must resolve visual motions up to 2000 deg/s. [O'Carroll et al. 1996] – and this using only a few drops of nectar a day.

1.1 Artificial Autonomous Systems

What applies to biological systems applies also to a large extent to any artificial autonomous system that behaves freely in a real-world[2] environment. When humankind started to build artificial autonomous systems, it was commonly accepted that such systems would become part of our everyday life by the year 2001. Numberless science-fiction stories and movies have encouraged visions of how such agents should behave and interfere with human society. Although many of these scenarios seem realistic and desirable, they are far from becoming reality in the near future. Briefly, we have a rather good sense of what these agents should be capable of, but we are not able to construct them yet. The (semi-)autonomous rover of NASA's recent Mars missions,[3] or demonstrations of artificial pets,[4] confirm that these fragile and slow state-of-the-art systems are not keeping up with our imagination.

Remarkably, our progress in creating artificial autonomous systems is substantially slower than the general technological advances in recent history. For example, digital microprocessors, our dominant computational technology, have exhibited an incredible development. The integration density literally exploded over the last few decades, and so did

[2]The term *real-world* is coined to follow an equivalent logic as the term *real-time*: a real-world environment does not really have to be the "real" world but has to capture its principal characteristics.

[3]*Pathfinder 1997, Mars Exploration Rovers 2004*: http://marsprogram.jpl.nasa.gov

[4]e.g. *AIBO* from SONY: http://www.sony.net/Products/aibo/

the density of computational power [Moore 1965]. By contrast, the vast majority of the predicted scenarios for robots have turned out to be hopelessly unrealistic and over-optimistic. Why?

In order to answer this question and to understand the limitations of traditional approaches, we should recall the basic problems faced by an autonomously behaving, cognitive system. By definition, such a system perceives, takes decisions, and plans actions on a cognitive level. In doing so, it expresses some degree of intelligence. Our goal-keeper knows exactly what he has to do in order to defend the free-kick: he has to concentrate on the ball in order to estimate its trajectory, and then move his body so that he can catch or deflect the ball. Although his reasoning and perception are cognitive, the immanent interaction between him and his environment is of a different, much more physical kind. Here, photons are hitting the retina, and muscle-force is being applied to the environment. Fortunately, the goalie is not directly aware of all the individual photons, nor is he in explicit control of all the individual muscles involved in performing a movement such as catching a ball. The goal-keeper has a nervous system, and one of its many functions is to instantiate a *transformation layer* between the environment and his cognitive mind. The brain reduces and preprocesses the huge amount of noisy sensory data, categorizes and extracts the relevant information, and translates it into a form that is accessible to cognitive reasoning (see Figure 1.1). This is the process of perception. In the process of action, a similar yet inverse transformation must take place. The rather global and unspecific cognitive decisions need to be resolved into a finely orchestrated ensemble of motor commands for the individual muscles that then interact with the environment. However, the process of action will not be addressed further in this book.

At an initial step perception requires sensory transduction. A sensory stage measures the physical properties of the environment and represents these measurements in a signal the

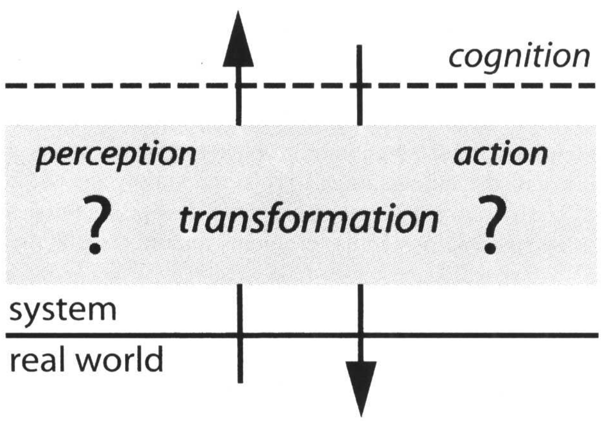

Figure 1.1 *Perception and action.*
Any cognitive autonomous system needs to transform the physical world through perception into a cognitive syntax and – vice versa – to transform cognitive language into action. The computational processes and their implementation involved in this transformation are little understood but are the key factor for the creation of efficient, artificial, autonomous agents.

rest of the system can process. It is, however, clear that sensory transduction is not the only transformation process of perception. Because if it were, the cognitive abilities would be completely overwhelmed with detailed information. As pointed out, an important purpose of perception is to reduce the raw sensory data and extract only the relevant information. This includes tasks such as object recognition, coordinate transformation, motion estimation, and so forth. Perception is the *interpretation* of sensory information with respect to the perceptual goal. The sensory stage is typically limited, and sensory information may be ambiguous and is usually corrupted by noise. Perception, however, must be robust to noise and resolve ambiguities when they occur. Sometimes, this includes the necessity to fill in missing information according to expectations, which can sometimes lead to wrong interpretations: most of us have experienced certainly one or more of the many examples of perceptual illusions.

Although not described in more detail at this point, perceptional processes often represent large computational problems that need to be solved in a small amount of time. It is clear that the efficient implementation of solutions to these tasks crucially determines the performance of the whole autonomous system. Traditional solutions to these computational problems almost exclusively rely on the digital computational architecture as outlined by von Neumann [1945].[5] Although solutions to all computable problems can be implemented in the von Neumann framework [Turing 1950], it is questionable that these implementations are equally efficient. For example, consider the simple operation of adding two analog variables: a digital implementation of addition requires the digitization of the two values, the subsequent storage of the two binary strings, and a register that finally performs the binary addition. Depending on the resolution, the electronic implementation can use up to several hundred transistors and require multiple processing cycles [Reyneri 2003]. In contrast, assuming that the two variables are represented by two electrical currents flowing in two wires, the same addition can be performed by simply connecting the two wires and relying on Kirchhoff's current law.

The von Neumann framework also favors a particular philosophy of computation. Due to its completely discrete nature, it forces solutions to be dissected into a large number of very small and sequential processing steps. While the framework is very successful in implementing clearly structured, exact mathematical problems, it is unclear if it is well suited to implement solutions for perceptual problems in autonomous systems. The computational framework and the computational problems simply do not seem to match: on the one hand the digital, sequential machinery only accepts defined states, and on the other hand the often ambiguous, perceptual problems require parallel processing of continuous measures.

It may be that digital, sequential computation is a valid concept for building autonomous artificial systems that are as powerful and intelligent as we imagine. It may be that we can make up for its inefficiency with the still rapidly growing advances in digital processor technology. However, I doubt it. But how amazing would the possibilities be if we could find and develop a more efficient implementation framework? There must be a different, more efficient way of solving such problems – and that's what this book is about. It aims to demonstrate another way of thinking of solutions to these problems and implementing

[5] In retrospect, it is remarkable that from the very beginning, John von Neumann referred to his idea of a computational device as an explanation and even a model of how biological neural networks process information.

them. And, in fact, the burden to prove that there are indeed other and much more efficient ways of computation has been carried by someone else – nature.

1.2 Neural Computation and Analog Integrated Circuits

Biological neural networks are examples of wonderfully engineered and efficient computational systems. When researchers first began to develop mathematical models for how nervous systems actually compute and process information, they very soon realized that one of the main reasons for the impressive computational power and efficiency of neural networks is the collective computation that takes place among their highly connected neurons. In one of the most influential and ground-breaking papers, which arguably initiated the field of computational neuroscience, McCulloch and Pitts [1943] proved that any finite logical expression can be realized by networks of very simple, binary computational units. This was, and still is, an impressive result because it demonstrated that computationally very limited processing units can perform very complex computations when connected together. Unfortunately, many researchers concluded therefore that the brain is nothing more than a big logical device – a digital computer. This is of course not the case because McCulloch and Pitts' model is not a good approximation of our brain, which they were well aware of at the time their work was published.

Another key feature of neuronal structures – which was neglected in McCulloch and Pitts' model – is that they make computational use of their intrinsic physical properties. Neural computation is physical computation. Neural systems do not have a centralized structure in which memory and hardware, algorithm and computational machinery, are physically separated. In neurons, the function is the architecture – and vice versa. While the bare-bone simple McCulloch and Pitts model approximates neurons to be binary and without any dynamics, real neurons follow the continuous dynamics of their physical properties and underlying chemical processes and are analog in many respects. Real neurons have a cell membrane with a capacitance that acts as a low-pass filter to the incoming signal through its dendrites, they have dendritic trees that non-linearly add signals from other neurons, and so forth. John Hopfield showed in his classical papers [Hopfield 1982, Hopfield 1984] that the dynamics of the model neurons in his networks are a crucial prerequisite to compute near-optimal solutions for hard optimization problems with recurrent neural networks [Hopfield and Tank 1985]. More importantly, these networks are very efficient, establishing the solution within a few characteristic time constants of an individual neuron. And they typically scale very favorably. Network structure and analog processing seem to be two key properties of nervous systems providing them with efficiency and computational power, but nonetheless two properties that digital computers typically do not share or exploit. Presumably, nervous systems are very well optimized to solve the kinds of computational problems that they have to solve to guarantee survival of their whole organism. So it seems very promising to reveal these optimal computational strategies, develop a methodology, and transfer it to technology in order to create efficient solutions for particular classes of computational problems.

It was Carver Mead who, inspired by the course "The Physics of Computation" he jointly taught with John Hopfield and Richard Feynman at Caltech in 1982, first proposed the idea of embodying neural computation in silicon *analog very large-scale integrated (aVLSI)* circuits, a technology which he initially advanced for the development of integrated digital