

Neural Interface and Atomic-force Microscope in CMOS Technology

A dissertation submitted to the

ETH ZURICH SWISS FEDERAL INSTITUTE OF
TECHNOLOGY

for the degree of

DOCTOR OF SCIENCE ETH

presented by

Sadik Hafizović

Dipl.-Ing. Microsystems, University of Freiburg, Germany

Born March 16th, 1977

Citizen of Germany

accepted on the recommendation of

Prof. Dr. Andreas Hierlemann

Prof. Dr. Sebastian Seung

Prof. Dr. Jan G. Korvink

Prof. Dr. Henry Baltes

Copyright © 2006 by Sadik Hafizovic, Physical Electronics Laboratory

All rights reserved. No part of this book may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior written permission of the copyright holder.

Cover page: Artistic rendering of the 10-cantilever AFM array scanning dried neurons. The first, left, short cantilever is a reference cantilever. Then, there are 10 scanning cantilevers, 4 of which can be simultaneously operated. The sample represents original data, see Figure 9.1 on page 78 and Figure 11.6 on page 103.

Published by:
Physical Electronics Laboratory
ETH Zurich
Wolfgang-Pauli-Strasse 16
8093 Zurich
Switzerland

Contents

1	Introduction	1
1.1	Structure of the Dissertation	1
1.2	Major Results	2
I	Neurochip	5
2	Introduction	7
2.1	Structure	8
3	Single-Chip Microelectronic System to Interface with Living Cells	9
3.1	Introduction	10
3.2	Methods	11
3.3	Results and Discussion	16
3.4	Conclusion	20
4	A CMOS-based Microelectrode Array for Interaction with Neuronal Cultures	23
4.1	Introduction	25
4.2	Materials and Methods	26
4.3	Results	39
4.4	Conclusion	45
4.5	Acknowledgements	46

5	Information Processing with Natural Neuronal Networks	47
5.1	Theoretical Neuroscience	47
5.2	Computing without States - Liquid State Machines	48
5.3	The Liquid State Machine Concept for Neurochips	50
5.4	Results	50
5.5	Discussion	57
6	Stimulation	61
6.1	Effective & Gentle Stimulation	61
6.2	Stimulation artifacts	62
7	Conclusion & Outlook	65
8	Neurochip Bibliography	67
II	Single-Chip Atomic Force Microscope	75
9	Introduction	77
9.1	Structure	77
10	CMOS Monolithic Mechatronic Microsystem for Surface Imaging and Force Response Studies	79
10.1	Introduction	80
10.2	Microcantilever Array	81
10.3	Circuit Design	83
10.4	Experimental Results	89
10.5	Conclusions	92
11	Single-Chip Mechatronic Microsystem for Surface Imaging and Force Response Studies	95
11.1	Introduction	96
11.2	Materials and Methods	96
11.3	Results and Discussion	101
12	Outlook	107

13 AFM Bibliography	109
Appendix	112
A Glossary	113
B Publications	115
C Acknowledgements	119
D Curriculum Vitae	121

Abstract

This thesis reports on the design, system integration, and experimental characterization of two different types of complementary-metal-oxide-semiconductor (CMOS)-based devices: a neurochip and an atomic-force microscope.

The neurochip reported on in this thesis is capable of bidirectionally communicating (stimulation and recording) with electrogenic cells such as neurons or cardiomyocytes and is targeted at investigating electrical signal propagation within cellular networks *in vitro*. The overall system consists of three major subunits: The core component is a 6.5-by-6.5 mm² CMOS chip, on top of which the cells are cultured. It features 128 bidirectional electrodes, each sampled at 20 kHz and 8 bit resolution. Second, there is a field-programmable-gate-array-based device, which provides chip control, event detection, data buffering and an USB interface capable of processing the 2.56 million samples delivered by the chip per second. The third element includes software that is running on a standard PC performing data capturing, processing, and visualization. Key challenges are posed by the low-noise circuitry (5.9 V_{rms} within 10 Hz - 100 kHz) that is required to reliably detect signals ranging from 1 mV_{pp} down to 40 μ V_{pp}, the amount of generated data (3.2 MB/s), and low-latency requirements (<2 ms between detection of spike and initiation of stimulation).

The conjunction of a highly integrated CMOS chip, which includes analog signal amplification and filtering stages, analog-to-digital converters, a digital-to-analog converter, stimulation buffers, temperature sensors, and a digital interface for data transmission, with a reconfigurable logic device (FPGA), improves the overall system performance. The advantages over current microelectrode arrays include the capability to stimulate (voltage stimulation, 8 bit, 60 kHz) spatio-temporal patterns on arbitrary sets of electrodes and the fast stimulation reset mechanism that allows to record neuronal signals on a stimulating electrode 5 ms after stimulation (instantaneously on all other electrodes). Also, the interconnect challenge that limits the size of currently available microelectrode arrays is overcome by on-chip multiplexing and A/D conversion.

Recordings from rat cortical neurons, chicken cortical neurons, rat brain slices, rat cardiomyocytes, and rat muscle cells have been attained. Additionally stimulation and recording measurements with rat neurons and rat cardiomyocytes have been successfully carried out and promising experiments on information processing with rat cortical neurons will be detailed.

The atomic-force microscope (AFM) is a single-chip (7 \times 10 mm²) AFM unit. It includes a fully integrated array of cantilevers, each of which has an individual actuation, detection, and control unit so that standard atomic force microscopy operations can be performed by means of the chip only without any external controller. The system offers drastically reduced overall size and costs as well as increased scanning speed and has been fabricated in standard CMOS technology with subsequent micromachining steps to form the cantilevers. Full integration of

microelectronic and micromechanical components on the same chip allows for the controlling and monitoring of all system functions. The system is monolithically integrated on a single chip and includes analog signal amplification and filtering stages with offset compensation, analog-to-digital converters, a powerful digital signal processor, and an on-chip digital interface for data transmission. The microsystem characterization evidenced a vertical resolution of <1 nm and a force resolution of <1 nN. The monolithic system represents a paradigm of a mechatronic microsystem that allows for precise and fully controlled mechanical manipulation in the nanoworld.

Zusammenfassung

Die vorliegende Dissertation beschreibt das Design, die Systemintegration und Funktion/Betrieb zweier CMOS-Mikrosysteme (complementary metal oxide semiconductor): (a) eines Neurochips und (b) eines Atomkraftmikroskops.

Der beschriebene Neurochip ist in der Lage mit elektrogenen Zellen, wie z.B. Neuronen oder Kardiomyozyten bidirektional (Stimulation und Aufzeichnung) zu kommunizieren. Das Ziel ist die Untersuchung der Ausbreitung elektrischer Signale in zellulären Netzwerken *in vitro*. Das Gesamtsystem besteht aus drei Einheiten: die Kernkomponente ist ein $6,5 \times 6,5 \text{ mm}^2$ großer Chip, auf dem die Zellen kultiviert werden. Der Chip umfasst 128 bidirektionale Elektroden, die mit je 20 kHz und 8 Bit Auflösung abgetastet werden. Weiter gibt es einen rekonfigurierbaren Logikbaustein (FPGA), der die Chipsteuerung, die Ereignisdetektion, die Kommunikationspuffer und die USB-Kommunikation bereitstellt. Das dritte Element ist die Software, die auf einem handelsüblichen Computer läuft und die Datenspeicherung, -prozessierung und -darstellung übernimmt. Die Anforderungen an das Chipdesign umfassen den gewünschten großen Rauschabstand, die große Menge an anfallenden Daten (3.2 MB/s) und die geforderte kurze Latenzzeit (1 ms).

Die digitale Verknüpfung des hochintegrierten CMOS-Chips, der die analogen Verstärker- und Filterstufen, die Analog-Digital- und Digital-Analog-Umsetzer, die Stimulationsverstärker und die Temperatursensoren beinhaltet, mit dem FPGA, erhöht die Systemleistung signifikant. Des Weiteren wird durch die digitale Schnittstelle die Anzahl der elektronischen Verbindungen reduziert und somit eine der größten Beschränkungen von derzeit verfügbaren Multielektrodenarrays eliminiert. Signale kortikaler Rattenneuronen, kortikaler Hühnerneuronen, Rattenhirnschnitten, Rattenkardiomyozyten und Rattenmuskelzellen wurden aufgenommen. Ausserdem wurde stimulierte Aktivität von kortikalen Rattenneuronen und Rattenkardiomyozyten aufgezeichnet.

Das beschriebene Atomkraftmikroskop (AFM) ist ein $7 \times 10 \text{ mm}^2$ großes Ein-Chip-AFM. Es besitzt ein vollintegriertes Array von Cantilevern, die jeweils über eine eigene Aktuations-, Detektions- und Kontrolleinheit verfügen, so dass für die Standard-AFM-Operationen keine externe Steuerungseinheit benötigt wird. Sowohl die Systemgröße als auch die Systemkosten sind dadurch drastisch reduziert, die Abtastgeschwindigkeit ist größer, und das System kann in einem Standard-CMOS-Prozess mit nachfolgenden Mikrobearbeitungsschritten für die Herstellung der Cantilever fabriziert werden. Die integrierten Schaltungen, welche analoge Verstärker- und Filterstufen mit Offsetabgleich, Analog-Digital-Umsetzer, einen leistungsfähigen digitalen Signalprozessor und eine digitale Schnittstelle beinhalten, sorgen für eine hohe Systemleistung. Die Charakterisierung des Mikrosystems ergab eine vertikale Auflösung von weniger als 1 nm und eine Kraftauflösung von weniger als 1 nN. Das monolithische AFM ist ein Beispiel für ein

mechatronisches Mikrosystem, das präzise und kontrolliert mechanische Manipulationen in der Nanowelt ermöglicht.

Chapter 1

Introduction

This thesis describes the design, system integration, and operation of two kinds of CMOS-based devices: a neurochip and an atomic force microscope (AFM). Both devices require similar CMOS design flows and feature comparable circuitry and system architectures. The respective transducers are formed during a post-CMOS in-house postprocessing. From this point on, the two projects differ considerably. While the neurochip electrodes are electrostatic transducers, the AFM cantilevers are mechanical transducers. Consequently, the postprocessing is very different for the two projects. Also the packaging does not resemble - both systems are highly optimized for a specific purpose.

For the results presented in this dissertation, it was necessary to recruit knowledge from a broad range of scientific and engineering disciplines. Such interdisciplinary work offers abundant possibilities for new concepts and tools, both for science and application, as the presented neurochip and the single-chip AFM demonstrates. This dissertation spans a wide range of activities - it includes digital circuit design, software engineering, informatics, micromachining and biology.

1.1 Structure of the Dissertation

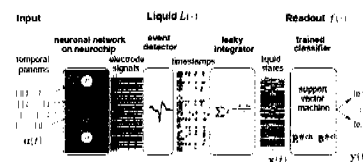
The core of this *paper dissertation* consists of a selection of four publications, two of which are related to the neurochip, and two of which describe the AFM. Hence, the thesis is structured into two parts:

1. Neurochip
2. Single-chip Atomic Force Microscope

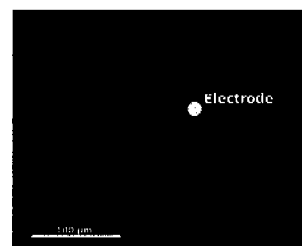
The neurochip activities required considerably more time and efforts so that the neurochip part is approximately twice as long as the AFM part. Both parts feature dedicated introductions, conclusions, and bibliographies.

1.2 Major Results

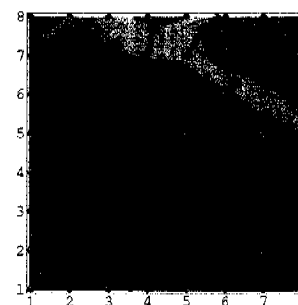
Information Processing with Natural Neuronal Networks. The first reported liquid state machine implemented with a natural neuronal network to our knowledge. The neuronal culture differentiates between two different stimulation patterns. A support vector machine has been used for the classification of the neuronal responses. (Journal Neuroscience Methods (Chapter 4 on page 23), Chapter 5 on page 47)



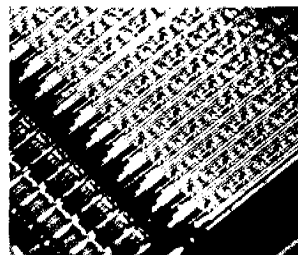
Neuronal Cultures. Recordings from rat cortical neurons, from rat acute brain slices and from chicken neurons. Rat and chicken neurons have been cultured and signals have been measured for longer than 2 months *in vitro*. (Chapter 3 on page 9)

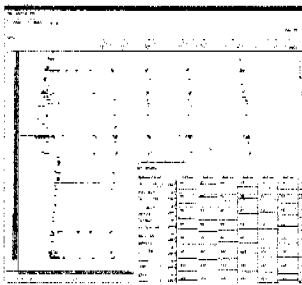


Cardiac Cultures. Dissociated chicken cardiomyocytes have been successfully cultured. The spatio-temporal evolution of the depolarization front was reconstructed for both, spontaneous and stimulated activity. (Proc. MEMS 2006, 3 on page 9)



128-channel CMOS Neurochip. A fully integrated microelectrode array with 128 bidirectional (stimulation and recording) electrodes and digital interface is presented. All electrophysiological measurements in this thesis have been obtained using this chip. (Chapter 3 on page 9)

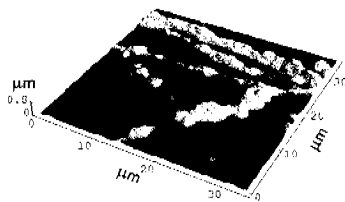




Neurotalker. Full-featured experimentation system for extracellular recordings comprising modular C++ software and FPGA-based event detection and USB2 communication. This software has enabled the presented experiments by providing real-time event detection, visualization, stimulation, as well as bidirectional feedback and recording experiments. (Chapter 4 on page 23)



Single-chip AFM. Milestone with regard to the complexity of monolithic microsystems. Analog signal conditioning with digital on-chip signal processor providing 16 million arithmetic operations per second for cantilever control. (JSSC 2005, 40(4), 951-959 (Chapter 10 on page 79), Proc. MEMS 2005)



Application of Single-chip AFM. Constant-force and force-distance measurements. A resolution of better than 1 nm and 1 nN is demonstrated. (PNAS 2004, 101(49), 17011-17015 (Chapter 11 on page 95))

Part I

Neurochip

Seite Leer /
Blank leaf

Chapter 2

Introduction

The work reported on in this part of the thesis has been performed in the framework of the INPRO project¹. The INPRO project is aimed at *INformation PROcessing through natural neuronal networks*. The underlying idea originates from fundamental differences of silicon-based information processing and information processing strategies as realized in nature. Table 2.1 summarizes some key figures of a bee brain, a human brain and a modern computer. It is apparent that the computational power of natural-neuron-based systems does not arise from their number of elements: a current microprocessor features more than 167 million transistors [Windeck, 2006], which is a hundred times the number of the neurons of a bee - neglecting that the microprocessor will still need memory.² Even when considering that a neuron arguably is a higher-order element than a transistor, it seems infeasible to attribute the computational performance to only the number of involved elements. Comparing the interconnectivity, which is a thousand times higher for the bee's brain in comparison to a silicon-based computer and even 10,000 times higher for a human brain, evidences that the reason for the outperformance is to be found in the architecture.

The implementation of a neural computation architecture in silicon [Mahowald and Douglas, 1991, Rasche and Douglas, 2000] is difficult since an enormous number and a high flexibility of the interconnections is required. Closer inspection reveals the restriction of silicon technologies to two dimensions as the origin of the limitations. A virtual implementation, in contrast will meet these two requirements, but since the implementing machine is a serial processor, the speed becomes the limiting issue. It is impressive to realize that even though implemented on such an inappropriate architecture like the universal processor, artificial neural networks have been so successful.

¹Funded by the Information Societies Technology (IST) Future and Emerging Technologies program of the European Union, contract number IST-2000-26463.

²The Intel Core Duo features 167 millions and 291 millions of transistors for the versions with 2 MB and 4 MB cache.

The thumb flip-book animation in the lower right corner is taken from the oral presentation at MEMS 2006, Istanbul. Hafizovic et al. [2006].

Table 2.1: Characteristic numbers of a modern computer, a bee brain and a human brain.

	Bee brain	Computer	Human brain
Speed	10^3 Hz	10^9 Hz	10^3 Hz
Elements	10^6 neurons	10^8 transistors	10^{12} neurons
Power Consumption	10^{-5} W	10^2 W	20 W
Connectivity (Fan Out)	10^3	10^1	10^4

2.1 Structure

The following part will consists of two papers,

1. *Single-Chip Microelectronic System to Interface with Living Cells* submitted to *Journal of Biosensors Biosystems* (Chapter 3) and
2. *System Integration of a CMOS-based Microelectrode Array for Interaction with Neuronal Cultures* submitted to the *Journal of Neuroscience Methods* (Chapter 4).

The first paper focuses on the MEA chip while the second covers the setup, software, event detection algorithm, and outlines a concept for information processing with natural neuronal networks. The following Chapter 5 is devoted to the topic of information processing with natural neuronal networks that was briefly addressed in paper 2 (Chapter 4). Finally, in Chapter 6, stimulation artifacts will be addressed, and in Chapter 7 the neurochip part of dissertation will be concluded.

For the applied neuronal and cardiac cell culture techniques, please refer to the related dissertations [Heer, 2005] and [Greve, 2006].

Chapter 3

Single-Chip Microelectronic System to Interface with Living Cells

F. Heer¹, S. Hafizovic¹, T. Ugniwenko², W. Franks¹, A. Blau², C. Ziegler²,
J.-C. Perriard³, and A. Hierlemann¹
in press at *Biosensors and Bioelectronics*¹

¹Physical Electronics Laboratory, ETH Zurich, Switzerland

²Department of Physics, University of Kaiserslautern, Germany

³Department of Biology, ETH Zurich, Switzerland

Abstract — A high degree of connectivity and the coordinated electrical activity of neural cells or networks are believed to be the reason that the brain is capable of highly sophisticated information processing. Likewise, the effectiveness of an animal heart largely depends on such coordinated cell activity. To advance our understanding of these complex biological systems, high spatiotemporal resolution techniques to monitor the cell electrical activity and an ideally seamless interaction between cells and recording devices are desired.

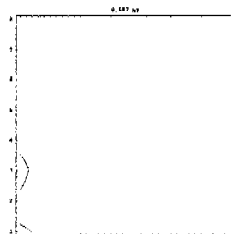
Here we present a monolithic microsystem in CMOS (complementary metal oxide semiconductor) technology that provides bidirectional communication (stimulation and recording) between standard electronics technology and cultured electrogenic cells. The microchip can be directly used as a substrate for cell culturing, it features circuitry units per electrode for stimulation and immediate cell signal treatment, and it provides on-chip signal transformation as well as a digital interface so that a very fast, almost real-time interaction (2 ms loop time from event recognition to, e.g., a defined stimulation) is possible at remarkable signal quality. The corresponding spontaneous and stimulated electrical activity recordings with neuronal and cardiac cell cultures will be presented.

¹Author contributions are:

Heer, Hafizovic, Hierlemann: neurochip

Ugniwenko, Blau, Ziegler: neuronal cultures

Perriard: cardiomyocyte cultures



The system can be used to, e.g., study the development of neural networks, to reveal the effects of neuronal plasticity and to study cellular or network activity in response to pharmacological treatments.

3.1 Introduction

Methods to directly measure electrical activity of cultured electrogenic cells like cardiomyocytes (heart cells) or neurons include two fundamentally different techniques:

1. transmembrane measurements by inserting one of the electrodes into the cell, the so-called *patch clamp* technique [Cole, 1949, Neher and Sakmann, 1976], and
2. extracellular recordings, e.g., by means of external microtransducers [De-Busschere and Kovacs, 2001, Fromherz, 2002, Jimbo and Robinson, 2000, Kleber and Rudy, 2003, Kovacs, 2003, Morefield et al., 2000, Olsson et al., 2005, Rutten, 2002, Pelt et al., 2004, Wise et al., 2004, Pancrazio et al., 2003, Marom and Shahaf, 2002].

Additionally, there are indirect methods like optical measurements using voltage-sensitive or fluorescent dyes [Baker et al., 2005, Obaid et al., 2004, Peterlin et al., 2000]. The patch clamp technique yields very accurate information on the electrophysiological properties of entire cells, or, alternatively, on currents flowing through single ion channels. However, it is an invasive method and is limited in the cell viability time (usually hours) and in the overall number of cells that can be simultaneously recorded from. For extracellular recordings, the cells are cultured directly on top of a transducer or an array of transducers (Figure 3.1). When an electrical activity or a so-called *action potential* occurs in a cell, ions flow across the cell membrane within milliseconds. When the cell is close enough to a transducer structure these moving ions generate an electric field or voltage, which either directly influences the open gate region of the field-effect transistor or can be recorded by the metallic microelectrode (Figure 3.1). Extracellular recordings are non-invasive (no puncturing of the cell membrane), which entails a potentially long measurement time, and microtransducer arrays offer multi-site measurement capabilities. For stimulation, voltage transients can be applied via the electrodes, which evoke a depolarization of the cell membrane and solicit subsequent electrical cell activity.

The electrodes are, in most cases, either metallic electrodes [Heuschkel et al., 2002, Jimbo et al., 2003, Maher et al., 1999, Stett et al., 2003b, Martinoia et al., 2004] or open-gate field-effect transistors [Jenkner et al., 2001, Voelker and Fromherz, 2005] on circuitless glass or silicon chips. For both transducers, CMOS (complementary metal oxide semiconductor)-based approaches to in vitro

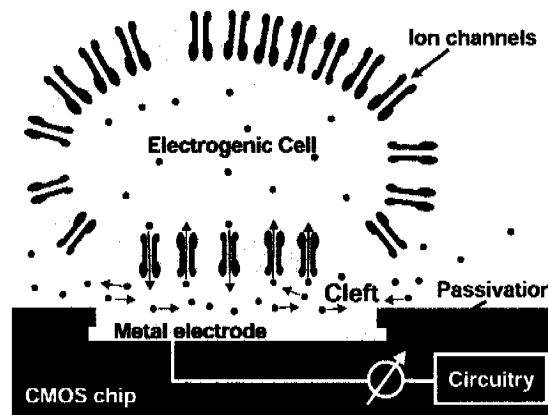


Figure 3.1: Extracellular measurements with metal electrodes: Moving ions in the electrode vicinity generate an electric field or voltage recorded by the microelectrode.

measurements featuring multi-addressing schemes and some basic signal treatment have recently emerged [Berdondini et al., 2005, DeBusschere and Kovacs, 2001, Eversmann et al., 2003a, Heer et al., 2004, Lambacher et al., 2004]. However, a CMOS-based array of electrodes, each electrode of which can be used for stimulation and recording, has not been reported on yet.

In this paper we present a single-chip 128-microelectrode array to bidirectionally connect to neuronal and cardiac networks in vitro. A micrograph of the overall system chip can be seen in Figure 3.2.a. The two main advantages in using integrated-circuit technology have been capitalized on: (a) signal quality; the signal is conditioned right at the electrode, and (b) connectivity; larger numbers of transducers can be addressed by on-chip multiplexing architectures. Figures 3.2.a,b and Figures 3.3.a,b show the chip architecture and its functional blocks including the electrode array with analog units repeated with each electrode, the analog-to-digital (A/D) and digital-to-analog (D/A) converters, the digital circuitry unit and the temperature sensor, which is used to maintain a constant temperature on the chip since the electrical cell activity, i.e., spike rates and amplitudes are strongly temperature-dependent.

3.2 Methods

3.2.1 Fabrication

The monolithic microsystem as presented in this paper relies on circular metal electrodes of 10 to 40 μm diameter at pitches between 50 and 250 μm (Figure 3.2.b). The key requirements for the electrode material are biocompatibility, process compatibility, low impedance and high charge-storage capacity. The electrode size and pitch can be arbitrarily selected since the electrode dimensions and locations are not defined by the CMOS process: only contact openings to the

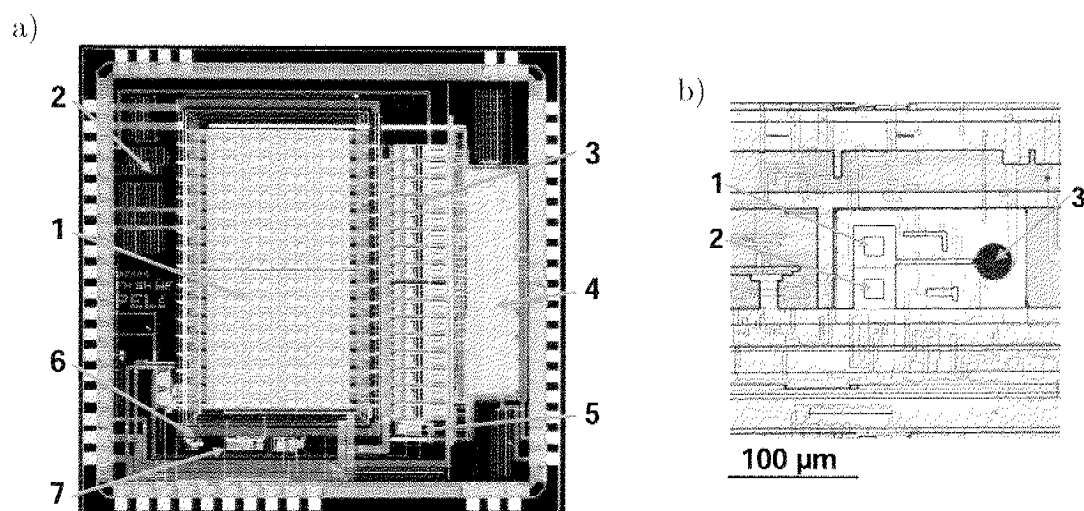


Figure 3.2: (a): Micrograph of the CMOS system chip ($6.5 \times 6.5 \text{ mm}^2$). The different components include: (1) 128-electrode array (8×16) including circuitry units per electrode; (2) space for on-chip platinum reference electrode; (3) sixteen analog-to-digital converters; (4) digital control and interface; (5) digital-to-analog converter for stimulation; (6) temperature sensor; (7) biasing circuitry.

(b): Micrograph of the electrode and the repeated circuitry unit for filtering, amplification, stimulation ($250 \times 250 \text{ µm}^2$); (1) stimulation contact; (2) recording contact; (3) platinum-black-covered electrode (30 µm diameter). The electrode is stimulation- and readout-capable since both contacts are electrically connected to the shifted electrode.

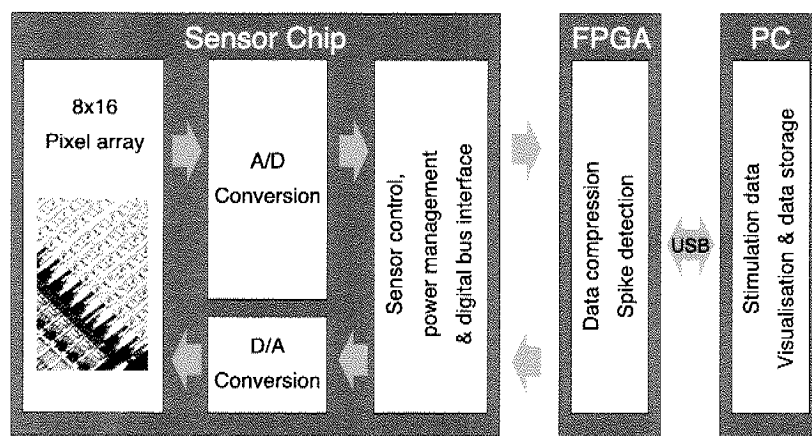


Figure 3.3: Overall setup for the CMOS microsystem: CMOS chip with 128 micro-electrodes, repeated analog units, analog-to-digital converters, a digital-to-analog converter, chip-control and power-management unit, and digital interface. FPGA board for data compression and event detection interfaced via USB to a computer for further data treatment and stimulus generation.

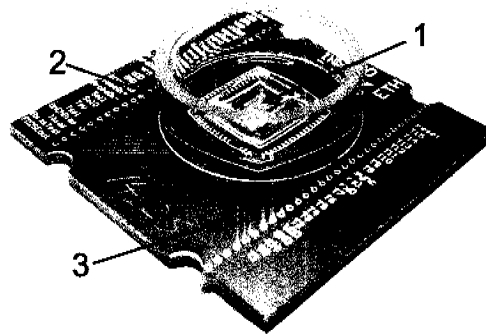


Figure 3.5: Photograph of the packaged chip ready for measurements; (1) CMOS chip; (2) glass ring; (3) printed circuit board.

The $6.5 \times 6.5 \text{ mm}^2$ chip has been fabricated using an industrial $0.6 \text{ }\mu\text{m}$ CMOS process at XFAB, Germany. After the CMOS process, a two-mask post-processing procedure has been used to fabricate the platinum electrodes. The processed chip has then been mounted and wire-bonded on a custom-designed printed-circuit board (PCB) (Figure 3.5). A water-resistant biocompatible epoxy (EPOTEK 302-3M, www.epotek.com) was used to encapsulate the bond wires and pads. A glass ring forms a bath that contains a suitable amount of cell medium. These post-processing and packaging techniques enable the microchips to withstand the cell culture environment for more than nine months so that a long-term interaction between state-of-the-art microelectronics and biology, i.e., cultured cells or cell networks is possible. Furthermore, we were able to clean and sterilize the chips after the termination of a cell culturing trial and to successfully record signals from re-used chips.

3.2.2 System description

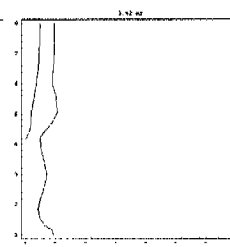
Electrical cell signals are in the kilohertz range so that 20,000 measurement values per second are acquired from each electrode for a faithful reproduction of the voltage transient. An oversampling at higher frequencies to achieve better temporal resolution is possible if not all electrodes in a row are selected for recording. Extracellular neuronal signals tend to be minute (tens to hundreds of microvolts) so that an analog circuitry repeating unit optimized for little area and low noise has been integrated with each electrode for immediate signal conditioning (Figures 3.3, 3.4). This circuitry unit features a fully differential architecture to reduce the effects of leakage currents, and electromagnetic interference. It incorporates a tunable band-pass filter and signal amplification stages. The high-pass filter (corner frequency tunable from 1 Hz to 1 kHz) removes offset and reduces drift and low-frequency noise of the metal electrodes in the aqueous solution, the subsequent low-pass filter (corner frequency tunable from 1 kHz to 50 kHz) limits the noise bandwidth to further improve the signal-to-noise ratio. Since each electrode is readout- and stimulation-capable, the repeating unit also includes

a buffer to ensure that each electrode can deliver the desired stimulation voltage (maximum output current 10 mA). For stimulation, typically a rectangular bipolar voltage signal of 100 μ s duration and up to 1 V amplitude is applied to a defined subset of electrodes, which then, in most cases, cannot be used for recording due to saturation of the amplifier electronics [Jimbo et al., 2003]. Here, we want to use the stimulating electrodes immediately after the stimulation event also for recording so that measures were taken to protect the sensitive readout circuitry during an electrical stimulus: The filters at each electrode can be individually reset to their operating point within 200 μ s to suppress the artifacts evoked by the voltage pulses. There is still a small artifact in the recording from the stimulating electrode, but the signal is back in the measurement range after approximately one millisecond and is back to baseline after three milliseconds as will be detailed later in this paper (context of Figure 3.10). During stimulation it is important to limit the voltage occurring at the electrode to obviate any electrolysis processes, which may damage the cells or the electrode. A variety of waveforms can be selected for stimulation since the signal can be generated with 60 kHz. The stimulation patterns and the selection of electrodes for stimulation can be varied at 20 kHz refreshing rate. Additionally, a small memory unit per electrode is provided to store its operation mode (stimulation, read, reset).

Each row of eight electrodes is multiplexed to a final amplification stage (Figure 3.3) (a total amplification of 1000 or 3000 can be selected) and then converted to the digital domain by an 8-bit successive-approximation analog-to-digital converter. A digital-to-analog converter translates the incoming digital stimulation signals into analog stimulation voltages.

The digital circuitry unit on chip controls the multiplexing, the electrode selection for stimulation, and the reset of single electrodes. Moreover, it controls the converters and it interfaces with the outside world (Figure 2a). The power consumption of the chip is 120 mW at 5 V supply, which does not lead to significant heating. A more detailed description of the CMOS circuit implementation can be found in [Heer et al., 2006a].

The overall setup is shown in Figure 3.3 and includes, besides the CMOS chip, a field-programmable gate array (FPGA) in conjunction with an universal serial bus (USB) 2.0 chip to manage the large data rates (3.2 MB/s out, 0.4 MB/s in). Input/output buffering and digital signal processing like averaging and event detection are implemented on the FPGA to reduce the data volume transmitted to the PC. This is an important asset in view of future larger arrays. The computer generates, on the one hand, the stimulation patterns and, on the other hand, serves as a data storage and visualization device. The time required to react upon the occurrence of a certain signal pattern in the cell culture with a defined stimulus is approximately 2 ms, which can be considered real-time so that the setup allows for fast closed-loop experiments [Wagenaar et al., 2005].



3.2.3 Cell culture preparation

The NRCs (neonatal rat cardiomyocytes) were dispersed from the ventricles of 1-3-day-old Sprague-Dawley rats by digestion with collagenase II (Worthington Biochemical Corp., U.S.A.) and pancreatin (Sigma-Aldrich, U.S.A.). To obtain cultures comprising more than 95% cardiomyocytes, the cell suspensions were separated on a discontinuous Percoll gradient (Sen et al., 1988). Cells were seeded in plating medium consisting of 68% DMEM (Amimed, Switzerland), 17% Medium M199 (Amimed), 10% horse serum (Life Technologies, USA), 5% fetal calf serum (Life Technologies), 4 mM glutamine (Amimed), and 1% penicillin-streptomycin (Amimed), compare. After 24 h the plating medium was exchanged for maintenance medium, consisting of 78% DMEM (Amimed), 20% Medium M1999 (Amimed), 1% horse serum (Life Technologies), 1% penicillin-streptomycin (Amimed), and 4 mM glutamine, (Amimed).

Hippocampal tissue was extracted from newborn Sprague-Dawley rats or embryonic chicken (*Gallus domesticus*) at embryonic day 10-12 (according to (Hamburger & Hamilton, 1951)) and dissociated by trypsinization (0.25% Trypsin/-EDTA) and gentle trituration. Cells were plated on laminin (10 g/ml) and Poly-L-Lysine (10 g/ml) coated CMOS chips at a density of 15000 cells/mm² and hold in Neurobasal medium (Gibco) containing B27 supplement (2%, Gibco), fatty acid supplement (0.1%, Sigma), lipid mixture (0.1%, Sigma), Alanine-Glutamine (2 mM, Gibco) and sodium pyruvate (1 mM, Sigma).

Prior to coating, the chips were prepared as follows: To optimize cellular adhesion, the chips were treated with oxygen plasma for 1 min at 40 W. The chips were then sterilized by immersing them in 70% ethanol for 1 min. After the ethanol was completely evaporated, the electrode area was rinsed several times with ultrapure water.

3.3 Results and Discussion

3.3.1 Recordings from neural and cardiac cultures

In a first experiment primary neonatal rat cardiomyocytes were cultured on the chips since these cells very quickly become electrically active and provide large signal amplitudes. Recordings from cardiomyocytes can be obtained after three days in culture, whereas dissociated neurons show spontaneous spiking only after about two to three weeks. For both cell types, the recording parameters included a bandpass filter range between 10 Hz and 5 kHz at a sampling frequency of 20 kHz to faithfully record the fast transient signals. The cardiomyocytes form a confluent layer on the chip and show spontaneous activity, a regular beating driven by pacemaker heart cells in the culture. Usually, the signals of several electrodes show the same beating rhythm so that so-called *field potentials* from

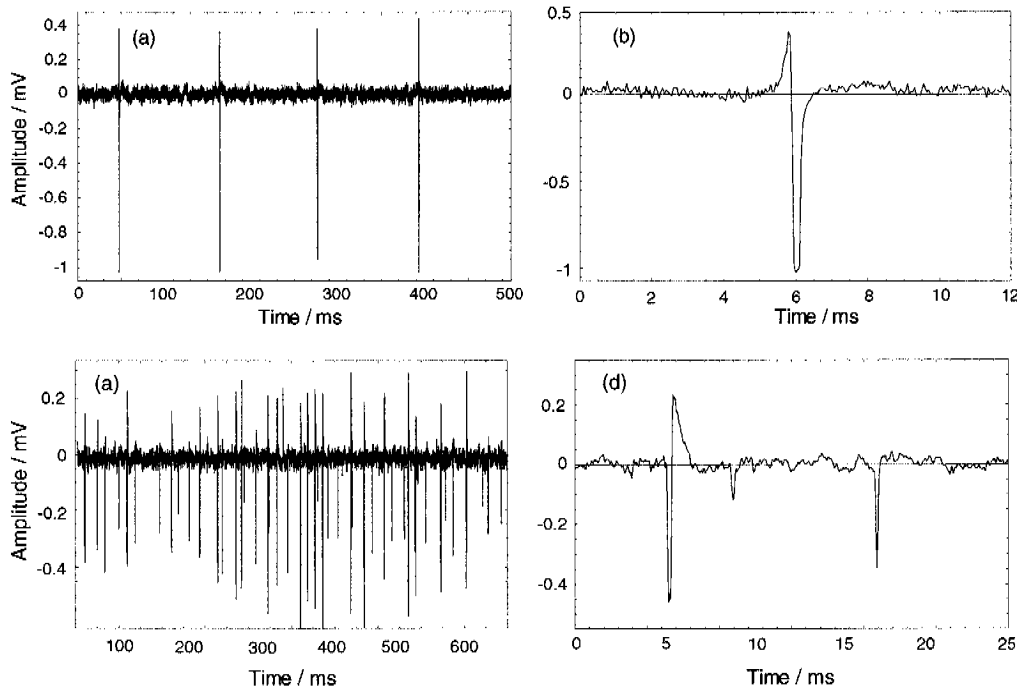
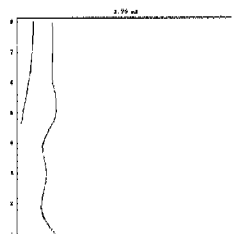


Figure 3.6: (a) Field potential recording of a confluent layer of regularly beating cardiomyocytes from neonatal rat (embryonic day 5) after phenylephrine dosage (100 $\mu\text{Mol/L}$) at 5 days in vitro; (b) close up of one individual spike; (c) activity recordings from dissociated chicken cortical tissue (embryonic day 10) after 27 days in vitro; (d) trace at higher temporal resolution showing several spikes.

patches of heart cells are recorded. Recordings of the spontaneously beating cells after 5 days in vitro are shown in Figures 3.6.a and 3.6.b. In this example, the cells have been activated by dosing phenylephrine and beat at a rate of about 10 Hz. The signals were recorded on 20- μm -diameter electrodes. The cardiomyocyte signals are rather large featuring amplitudes between 1 and 1.5 mV. The lower graph shows one of the spikes at an extended time scale.

In a next experiment, neural networks originating from dissociated cortical tissue of fertilized chicken eggs (*Gallus domesticus*) at embryonic day 9 (Hamburger & Hamilton, 1951) were grown on CMOS chips. These neurons survived more than 9 months on the CMOS chip, which suffices to possibly perform some kind of learning or information processing. Exemplary signals from spontaneously firing cells after 27 days in vitro are shown in Figures 3.6.c and 3.6.d. The noise level in these recordings was 17 μV_{RMS} at the given bandpass range and sampling frequency. The signals recorded from 20- μm -diameter electrodes showed rather large signals between 500 μV and 700 μV , the amplitudes of which are most probably a consequence of the immediate on-chip signal conditioning.



3.3.2 Stimulation of neural and cardiac cell cultures

Stimulation experiments have been carried out using again cardiomyocytes first (Figure 3.7). Half of the array (64 electrodes) was used for this purpose. The spontaneous regular beating of the cells every 1.2 seconds is shown in Figure 3.7.a, which is an electrode event plot versus time, i.e., each electrical signal on an electrode is represented as a dot. The stimulation included 20 bipolar pulses of ± 1 V at 1.5 kHz as can be seen in Figure 3.7.b, which shows the stimulation signal on an electrode at 750 μm distance from the stimulated electrodes (reset function was not active at this electrode to show the stimulation), and the evoked electrical activity after the stimulation. The cell culture reacts upon stimulation and adopts the stimulation signal as a new starting point for its spontaneous

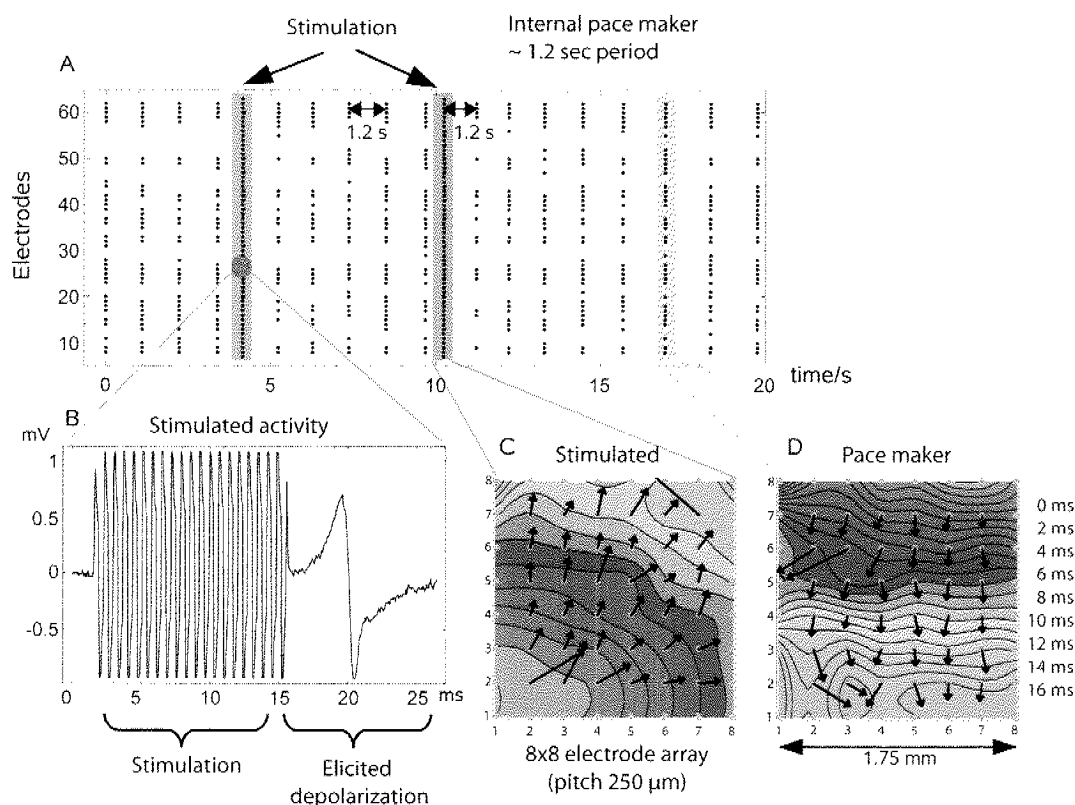


Figure 3.7: Stimulation and recording on a neonatal rat cardiomyocyte culture (embryonic day 5) after 3 days in vitro. (a) Electrode event plot versus time (each electrode signal represented as a dot) showing spontaneous regular beating of the cells every 1.2 seconds and the stimulations; (b) stimulation (20 bipolar pulses of ± 1 V at 1.5 kHz) and elicited cell activity on an electrode at 750 μm distance from stimulated electrodes (reset was not activated to show the stimulation); (c) isochronous map of a stimulated electrical wave propagating from the lower left (stimulation on electrodes 1-8); the times are encoded in different colors (d) isochronous map of a spontaneous electrical wave propagating from a pacemaker region at the upper right corner. The propagation of the stimulated wave (12-14 ms across the array) is somewhat faster than that of the spontaneous wave (16-18 ms).

activity, as can be seen even more clearly for the second stimulation. Figures 3.7.c and 3.7.d show the propagation of spontaneous and triggered electrical waves over the electrode area. The pacemaker cells that produce spontaneous activity are in the upper right corner, and the wave propagates from there within about 16 ms over the electrode array (Figure 3.7.d). For stimulation, electrodes in the lower left corner were selected, which caused a signal wave in the opposite direction.

Finally, stimulation experiments using cortical neurons from the rat brain have been carried out. Figure 3.9 shows an example of a successful excitation of spikes recorded at 250 μm distance from the stimulation site. A single bipolar stimulation pulse of ± 800 mV and 50 μs duration was used to stimulate the cells in this example. The inset shows the post-stimulus time histogram of this channel that includes the results of 142 stimulation pulses. The first event or neuronal signal generally occurs between two and three milliseconds after the stimulation pulse with a very high probability of 96%.

3.3.3 Stimulation artifact suppression

Figure 3.10 shows the efficiency of the on-chip reset function for stimulation artifact suppression. Bipolar pulses (220 μs overall duration) of different amplitudes (± 0.15 V and ± 1 V) in saline solution have been applied, while the reset has been operational (trace (a) ± 0.15 V; trace (b) ± 1 V) or not (trace (c) ± 0.15 V, trace (d) ± 1 V). The recordings from the stimulation electrode are shown. The readout filter for the red and green trace was reset 50 μs before the stimulation pulse was applied and kept until 50 μs after the stimulation pulse was finished.

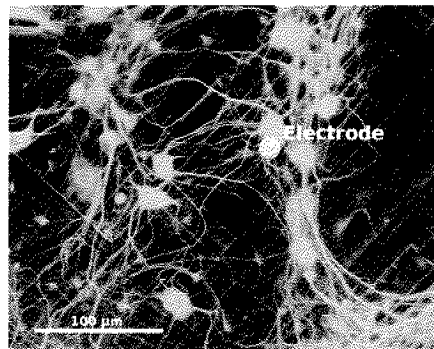
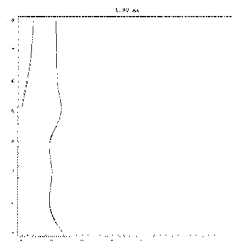


Figure 3.8: Fluorescence image of a neural network as grown on the CMOS chip; Neurons originated from the rat hippocampal tissue of newborn Sprague-Dawley rats. Neural networks were cultured for 17 days in supplemented serum-free Neurobasal medium. MAP2 (microtubuli-associated protein-2) immunostaining of the neurons was visualized using an FITC-conjugated secondary antibody (green). The structure of the CMOS chip surface is visible in the background. A color version is printed on page 41.



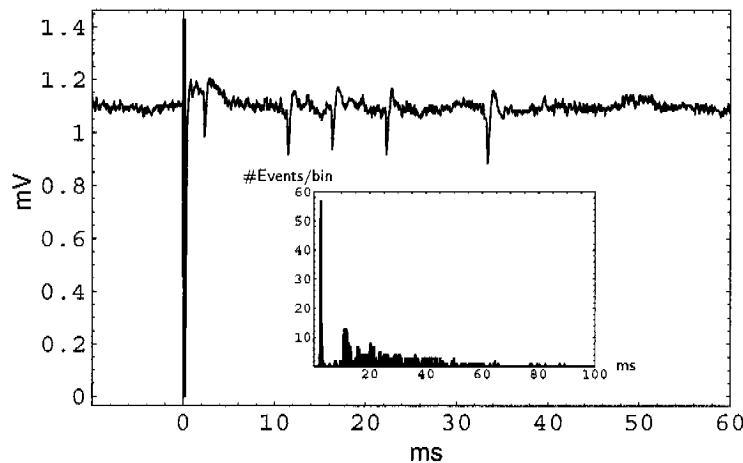


Figure 3.9: Successful excitation of spikes in cortical neurons from rat brain at 250 μm distance from the stimulation site. A single bipolar stimulation pulse of ± 800 mV and 50 μs duration was used to stimulate the cells in this example. The inset shows the post-stimulus time histogram of this channel based on 142 stimulation pulses.

Without reset, it takes 20 to 100 ms (depending on the stimulation amplitude) for the recording circuitry to return to the measurement range. The inset shows a close-up of the initial 20 ms at better temporal resolution. With reset, it takes less than 5 ms for the recording circuitry to return to the measurement range. It is worth mentioning that a stimulation sequence always ends with a stimulation of a value close to the equilibrium potential of the electrode so that the electrode quickly returns to its equilibrium potential after stimulation.

3.4 Conclusion

The presented single-chip system fabricated in industrial CMOS technology combined with post-CMOS processing is very compact, and there is no electrical shielding needed for its operation. The system enables high-spatiotemporal-resolution, bidirectional interaction with electrogenic cells and will find numerous applications in, e.g., neuroscience and pharmacology [DeBusschere and Kovacs, 2001, Fromherz, 2002, Jimbo and Robinson, 2000, Kleber and Rudy, 2003, Kovacs, 2003, Marom and Shahaf, 2002, Morefield et al., 2000, Olsson et al., 2005, Rutten, 2002, Pelt et al., 2004, Wise et al., 2004, Pancrazio et al., 2003]. Moreover, it may be used as a point-of-care device to, e.g., test potential medication on patient cells before administration, or it may become a prototype device to combine the parallel information processing characteristics of natural neurons with the more serial ones of microelectronics to realize new ways of *bioelectronic* information processing.

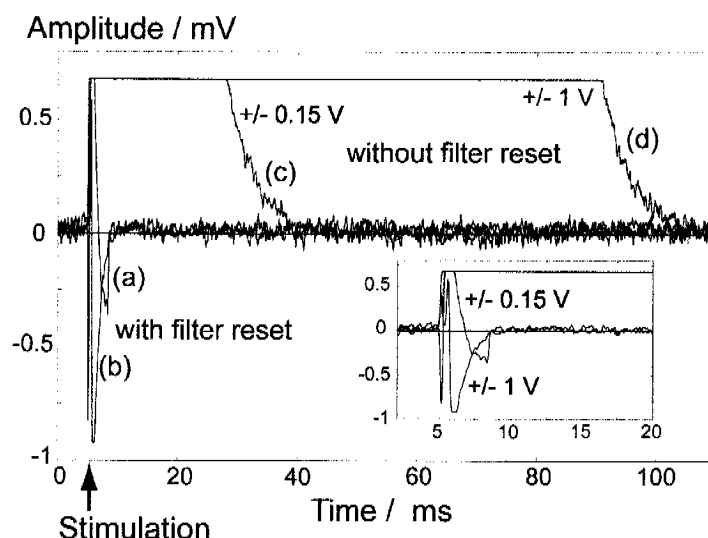
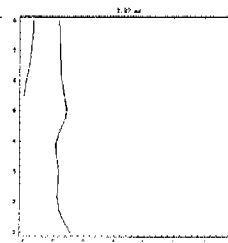


Figure 3.10: Efficiency of the on-chip reset function for stimulation. Bipolar pulses (220 μ s overall duration) of different amplitudes (± 0.15 V and ± 1 V) in saline solution have been applied, and the reset has been operational (trace (a) ± 0.15 V; trace (b) ± 1 V) or not (trace (c) ± 0.15 V, trace (d) ± 1 V). Without reset, it takes 20 to 100 ms for the recording circuitry to return to the measurement range. The inset shows a close-up of the initial 20 ms at better temporal resolution.

Acknowledgements

The authors are greatly indebted to staff of the Physical Electronics Laboratory at ETH Zurich involved in the biochip and biomicrosensor development, Urs Frey, Frauke Greve, and Kay-Uwe Kirstein. Moreover, the authors are grateful to Prof. Henry Baltes, Physical Electronics Laboratory, ETH Zurich for sharing laboratory resources and for his ongoing stimulating interest in their work. The authors also want to thank Evelyne Perriard at ETH Zurich and Simone Riedel at the University of Kaiserslautern for assistance with the cell preparations. Funding has been generously provided by the Information Societies Technology (IST) Future and Emerging Technologies program of the European Union, and the Swiss Bundesamt für Bildung und Wissenschaft (BBW) under contract number IST-2000-26463.



Chapter 4

A CMOS-based Microelectrode Array for Interaction with Neuronal Cultures

Sadik Hafizovic¹, Flavio Heer¹, Tanja Ugniwenko², Axel Blau²,
Christiane Ziegler², and Andreas Hierlemann¹,
submitted to *Journal of Neuroscience Methods*

¹Physical Electronics Laboratory, ETH Zurich, Switzerland

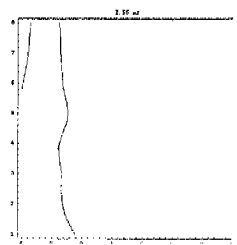
²Department of Physics, University of Kaiserslautern, Germany

Abstract — We report on the system integration of a CMOS chip that is capable of bidirectionally communicating (stimulation and recording) with electrogenic cells such as neurons or cardiomyocytes and that is targeted at investigating electrical signal propagation within cellular networks *in vitro*.

The overall system consists of three major subunits: First, the core component is a 6.5-by-6.5 mm² CMOS chip, on top of which the cells are cultured. It features 128 bidirectional electrodes, each equipped with dedicated analog filters and amplification stages and a stimulation buffer. The electrodes are sampled at 20 kHz with 8 bit resolution. The measured input-referred circuitry noise is 5.9 μ V root mean square (10 Hz to 100 kHz) which allows to reliably detect the cell signals ranging from 1 mV_{pp} down to 40 μ V_{pp}. Additionally, temperature sensors, a digital-to-analog converter for stimulation, and a digital interface for data transmission are integrated. Second, there is a reconfigurable logic device, which provides chip control, event detection, data buffering and an USB interface capable of processing the 2.56 million samples delivered by the CMOS chip per second. The third element includes software that is running on a standard PC performing data capturing, processing, and visualization.

Experiments involving the stimulation of neurons with two different spatio-temporal patterns and the recording of the triggered spiking activity have been carried out. The response patterns have been successfully classified (83% correct) with respect to the different stimulation patterns.

The advantages over current microelectrode arrays, as has been demonstrated in



the experiments, include the capability to stimulate (voltage stimulation, 8 bit, 60 kHz) spatio-temporal patterns on arbitrary sets of electrodes and the fast stimulation reset mechanism that allows to record neuronal signals on a stimulating electrode 5 ms after stimulation (instantaneously on all other electrodes). Other advantages of the overall system include the small number of needed electrical connections due to the digital interface and the short latency time that allows to initiate a stimulation less than 2 ms after the detection of an action potential in closed-loop configurations.

4.1 Introduction

Microelectrode arrays (MEA) [Morin et al., 2005, Potter, 2001] are used to extracellularly measure the induced voltage on an electrode underneath a cell upon the occurrence of an action potential [Hodgkin and Huxley, 1952]. The obtained electrode signal amplitudes typically range between 100 μV and 1 mV, depending on the cell type. With increasing cell-electrode distance or less direct cell-electrode contact, signal amplitudes may become arbitrarily small [Claverol-Tinture and Pine, 2002].

MEAs have become an important tool for non-invasive recording in the fields of neuroscience and biosensing [Morin et al., 2005, Marom and Shahaf, 2002, Baumann et al., 2004]. Artificial neural networks on MEAs can potentially be trained to perform, e.g., pattern recognition tasks [Marom and Shahaf, 2002, Wagenaar et al., 2006b]. Additionally, the ability to monitor the electrophysiological response of a cell, or cell culture to pharmacological agents facilitates applications in the area of pharascreeing [Stett et al., 2003b]. Previously described MEAs

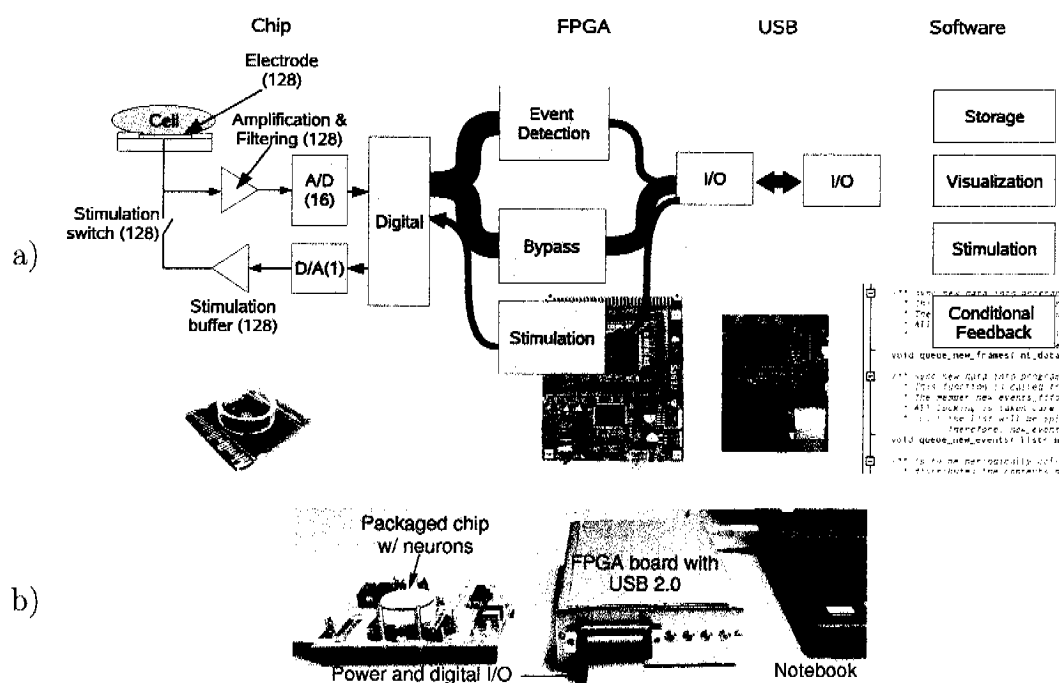
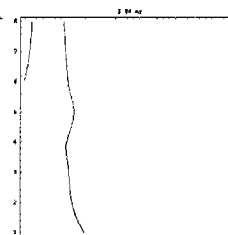


Figure 4.1: (a) System overview and data flow: Recorded data propagate from the left to the right (chip to PC) while stimulation data propagate from the right to the left (PC to chip). The outermost left block labeled “Chip”, includes components in multiple realizations: numbers in brackets indicate the number of realizations per chip, e.g. 128 electrodes or 16 A/D converters. For the sake of clarity, the analog signal processing chain is highly simplified. A chip circuitry schematic is given in Figure 4.3 and a simplified FPGA schematic is given Figure 4.5;

(b) Photo of the entire system setup: The PCB provides a stabilized power supply and reference voltages for the CMOS chip.



typically are either passive devices [Maher et al., 1999, Jimbo et al., 2003, James et al., 2004, Buitenweg et al., 2002c, MCS, Gross et al., 1977], or rely on integrated multiplexing circuits combined with off-chip electronics for system control and signal conditioning [Eversmann et al., 2003a, Baumann et al., 2004, Berdoncini et al., 2005]. The passive devices are usually robust and deliver good signal quality but are limited in the number of electrodes by the number of electrical interconnects. MEAs with integrated multiplexing but without on-chip filtering feature more electrodes but often suffer from lower signal quality, mainly due to the lack of aliasing filters.

The CMOS-MEA presented here is highly integrated and features 128 bidirectional electrodes, each of which has dedicated signal conditioning circuitry and is sampled at 20 kHz and 8 bit resolution. Any subset of electrodes may be chosen for stimulation of the cell culture. After stimulation has been terminated, the read-out is immediately reestablished due to an effective artifact suppression technique (see details in section 4.2.2.2 and Figure 4.7). The digital on-chip circuitry combined with external reconfigurable logic (field-programmable gate array, FPGA) for event detection entails a very short response time so that the system may be considered real-time with respect to neural cell signaling. A stimulation pattern can be triggered within 2 ms after detection of an event (latency < 2 ms).

4.2 Materials and Methods

The system includes three custom-engineered components: (1) the MEA ASIC, (2) the FPGA and (3) the software. Figure 4.1 depicts the three components and their interfacing. Major challenges arise from the operating conditions (incubator: 95% rel. humidity, 37°C), the requested biocompatibility of chip and package, the large data volumes (3.2 MB/s) that have to be handled, the need for low-noise analog amplifiers and filters, and the low-latency requirements (2 ms).

4.2.1 Chip Design

The micrograph of the 6.5 by 6.5 mm² chip in Figure 4.2 (left) shows three main areas, which are clearly distinguishable: the 128-electrode array, the column of 16 A/D-converters, and the digital core. The design is modular in that the same electrode unit is repeated 128 times, and in that each row of eight electrode units includes an A/D converter, so that there are 16 analog-to-digital (A/D) converters in total. Considerable space between the electrodes and the bondpads is needed to allow for an effective sealing of the electrical interconnects off chip.

The raw electrode signal of 0.1 to 1 mV_{pp} is conditioned in a circuitry block at each electrode, then multiplexed and further amplified before it undergoes the

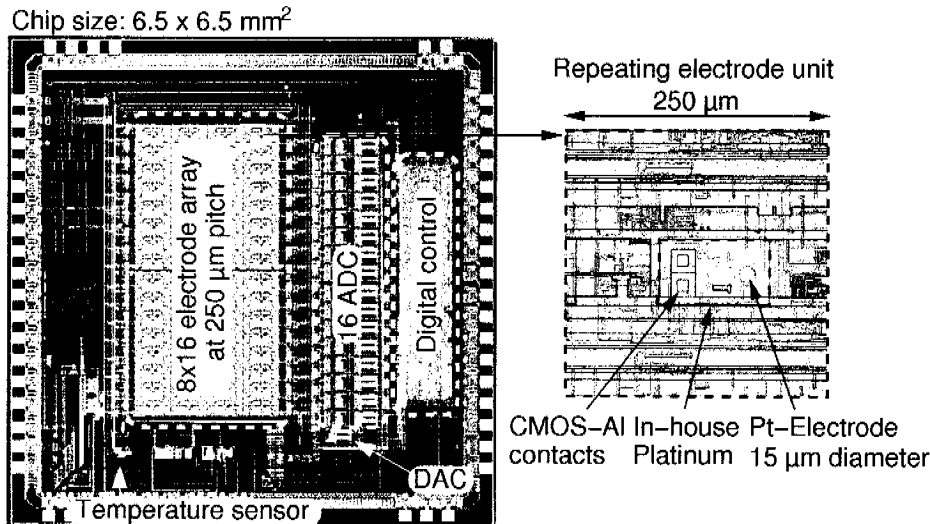


Figure 4.2: Micrograph of the electrode array chip and close-up of one electrode repeating unit. Left: The chip features an 8-by-16 electrode array in the center, and 16 A/D-converters and the digital block at the right. Right: Close-up of the 128-times-repeated circuitry unit.

20-kHz, 8-bit A/D-conversion. The total amplification is selectable, either 1,000 or 3,000-fold.

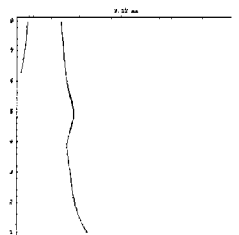
The combination of sophisticated analog signal processing with a digital interface renders the overall multi-electrode array system very compact. Figure 4.1 (right) shows all needed components: a laptop computer, an USB-powered FPGA card, and a simple PCB to provide reference voltages and power supply stabilization.

Electrogenic cells are very sensitive to temperature so that temperature changes may alter the cell activity or even lead to cell death. An on-chip temperature sensor is hence provided to monitor the cell-area temperature.

4.2.1.1 Electrode Array / Analog Circuitry

Each of the 250-by-250 μm^2 electrode circuitry repeating units in the array includes a high-pass filter with 20 dB gain, a low-pass filter, a 30-dB amplifier, a stimulation buffer, and digital circuits for 2-bit mode storage as shown in Figure 4.3. Electromagnetic interferences are reduced and an effective power supply rejection is realized by means of a fully differential architecture. The design goal for the input-equivalent noise of the circuitry is determined by the thermal noise of the platinum electrodes (section 4.2.4) in physiological saline solution, which has been measured to be 3.19 μV_{rms} for $10 \times 10 \mu\text{m}^2$ and 2.16 μV_{rms} for $40 \times 40 \mu\text{m}^2$ electrodes within a band from 10 Hz to 100 kHz.

In comparison to other CMOS electrode arrays [Eversmann et al., 2003a, Kovacs, 2003], important advantages arise from the modular architecture with buffers and filters implemented at each electrode:



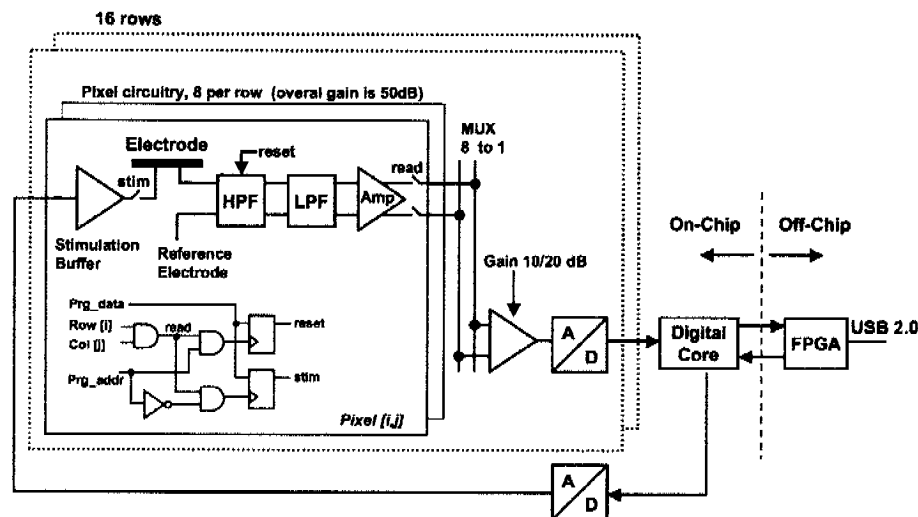


Figure 4.3: Schematic of the on-chip electronics architecture and its components. The stacked frames indicate that these subunits are repeated for each electrode or each row. The chip also includes a temperature sensor. HPF denotes high-pass filter, LPF low-pass filter. A concise description of the on-chip analog circuitry is provided in [Heer et al., 2006a].

(i) The high-pass filter removes offset and slow drifts of the biochemical signals and, thereby, allows for immediate signal amplification; (ii) the low-pass filter limits the noise bandwidth and acts as an anti-aliasing filter for the subsequent A/D-conversion; (iii) the signal is amplified and filtered in close vicinity of the electrodes, which makes the design less sensitive to noise and distortion picked up along connection lines; (iv) a stimulation buffer makes the stimulation signal independent of the number of activated electrodes. Additionally, the high-pass filter has a reset in order to ensure operability of the sensitive recording electronics also immediately after stimulation.

A limitation of the number of electrodes arises from the chip area needed for the individual electrode circuitry units. A MEA based on the electrode circuitry unit presented here ($250 \times 250 \mu\text{m}^2$) with 1024 electrodes would consume a total chip area of $8 \times 8 \text{ mm}^2$ and could be readily implemented. However, electrode numbers as high as 16,384 as reported in [Eversmann et al., 2003a] cannot be realized with this design.

Stimulation Desired stimulation features include:

- Possibility to stimulate any subset of electrodes
- Flexibility in the stimulation waveform
- Current and voltage stimulation capability

- Capability to do a fast switching from stimulation to recording even on the same electrode

The real-time all-channel stimulator by Wagenaar and Potter [2004], based on discrete off-chip components, offers these features except for the last: the time during which an electrode cannot be used for recording after stimulation was reported to be longer than 40 ms. The presented monolithic CMOS implementation meets all criteria except for the possibility to do current stimulation. Stimulation of neuronal cells can be performed by means of current or voltage pulses. Intracellular stimulation during patch clamp measurements is usually done in the current mode, since the stimulation of neurons in their natural environment also occurs through current inputs to their dendritic tree. For extracellular stimulation both, voltage stimulation [Taketani and Baudry, 2006, Bonifazi and Fromherz, 2002, Gross et al., 1993] and current stimulation [Taketani and Baudry, 2006, McIntyre and Grill, 2002, Jimbo and Kawana, 1992, Buitengeweg et al., 2002b] methods have been widely used. A detailed description of stimulation methods for cell cultures with microelectrode arrays is given in [Wagenaar et al., 2004].

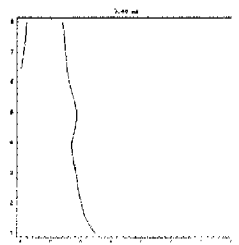
Biphasic voltage-controlled pulses with the first part of the pulse having a positive voltage sign constitute effective stimuli, since the sharp downward voltage transient between the positive and the negative phase corresponds to a strong negative current pulse. The pulse amplitude and its duration determine the stimulus efficacy [Wagenaar and Potter, 2004, McIntyre and Grill, 2002, Buitengeweg et al., 2002b]. Another important parameter is the electrode impedance, since it determines the current for a given voltage transient and the voltage on the electrode for a given current transient. Impedance values for the type of platinum electrodes used here have been measured and published in [Heer et al., 2004].

An 8-bit flash D/A-converter has been implemented for stimulation. An arbitrary stimulation signal with a sampling frequency of 60 kHz can be applied to any subset of electrodes. Since stimulation voltages typically are in the range of 1 V, and since recorded signals are below 1 mV, any stimulation will saturate the respective signal conditioning chain. To allow for immediate read-out after stimulation, the stimulating electrodes can be reset ("reset" in Figure 4.3) to their operating points. All the on-chip circuitry has been described in [Heer et al., 2006a] in great detail.

4.2.1.2 Digital On-chip Circuitry

The large data volumes and the desired low latency time require high-performance communication and signal processing capabilities. Digital logic components distributed on the MEA ASIC and on the FPGA bridge the A/D-converters to the PC.

The on-chip digital logic runs at 1.6 MHz and 5 V and serves two purposes: Firstly, it performs control tasks like multiplexing, electrode selection for stimulation, reset of single electrodes, and it contains the successive-approximation



registers of the A/D converters; Secondly, it provides the chip interface to the FPGA. In the architecture shown here, this digital interface uses 16 lines for the data readout, 2 lines for stimulation and control data, and additionally, a clock, a reset and a *frame sync* line. All signals are synchronous to the 1.6 MHz chip clock.

4.2.2 FPGA

To manage the data output rate of 3.2 MB/s and the input rate of 0.4 MB/s, an FPGA running at 48 MHz in conjunction with an USB 2.0 interface chip and 1 MB SRAM CESYS GmbH has been utilized. I/O buffering and digital signal processing like lowpass filtering and event detection are implemented on the FPGA to reduce the data volume transmitted to the PC.

4.2.2.1 Event Detection

Event detection is mandatory since data volumes are beyond the bandwidth of a human observer. In [Lewicki, 1998] a review of common approaches is given. The approach as presented in this paper is realized as a two-stage processing:

1. During *event detection*, events are isolated from the continuous flow of the data. In this context an event is a segment of recorded data that potentially contains a single spike, multiple (possibly overlapping) spikes, a spike train or barrage.
2. During *spike discrimination*, events are further processed to isolate spikes, to perform a unit separation and to resolve overlaps.

The benefits of performing event detection on the FPGA as opposed to a PC are twofold. First, the PC is relieved from approximately 32 million arithmetic operations per second. Second, the load on the I/O system is decimated, which is a prerequisite for a future realization of larger arrays. The freed PC resources can be utilized for more complex tasks like on-line spike sorting, which are hard to perform in hardware. Additionally, the latency time of the whole system is reduced.

Algorithm Two statistical variables per electrode are needed for event detection: the moving average of the signal, \bar{x}_i , and the moving variance, σ_i^2 , where $i = 0 \dots 127$ denotes the electrode number. As illustrated in Figure 4.4, events are detected by comparing the absolute value of the highpass-filtered signal, $|x_i - \bar{x}_i|$, with four times the standard deviation, σ_i :

$$(x_i - \bar{x}_i)^2 > 4^2 \sigma_i^2. \quad (4.1)$$

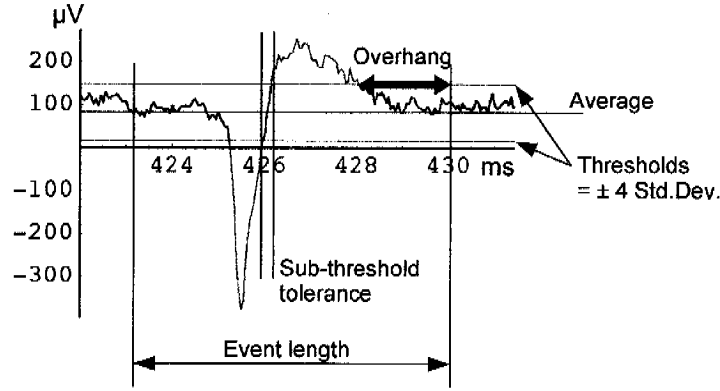


Figure 4.4: Event detection exemplified using a chicken cortical neuron spike. The central strip indicates the 4σ -region. The onset of an event is registered when the signal leaves the 4σ -region for longer than a defined minimum time (4 samples = 0.2 ms). The end of an event is registered when an *overhang*-long sequence of sub-threshold values is found, where *overhang* can be defined to, e.g., 16 samples.

The mean signal and the variance are continuously updated by numeric infinite-input-response (IIR) lowpass filters with:

$$\bar{x}_i(k) = (1 - \varepsilon_{\bar{x}}) \times \bar{x}_i(k-1) + \varepsilon_{\bar{x}} \times x_i(k)$$

and

$$\sigma_i^2(k) = (1 - \varepsilon_{\sigma}) \times \sigma_i^2(k-1) + \varepsilon_{\sigma} \times (x_i(k) - \bar{x}_i(k))^2,$$

respectively. Using σ^2 instead of σ circumvents square-root calculation. Sensible coefficients are $\varepsilon \in \{2^{-n} | n \in \{6, 7, 8, 9\}\}$, which allows to use shift operations instead of multiplications and to keep the requirements concerning the fix-point precision manageable. The required fix-point accumulator width is determined by the input/output width (8 bit for \bar{x}_i , 6 bit for σ_i^2) plus n post-decimal-point digits. Hence, higher n -values require higher fix-point precision and, consequently, more storage space. The corresponding time constants, τ , with $\varepsilon = 2^{-n}$ are given by

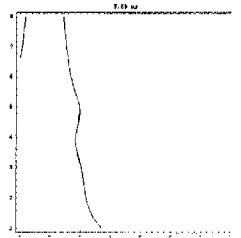
$$\tau = -\frac{1}{f_{\text{sample}}} \frac{1}{\ln(1 - 2^{-n})}, \quad (4.2)$$

with f_{sample} being the sampling frequency.

For $\varepsilon_{\bar{x}} = 2^{-6}$ and $\varepsilon_{\sigma} = 2^{-7}$, the respective time constants are 3.17 ms and 6.37 ms. By using only every second value to update \bar{x}_i and σ_i^2 , f_{sample} is effectively cut in half (10 kHz instead of 20 kHz). This way, the time constants are doubled to $\tau_{\bar{x}} = 6.35$ ms and $\tau_{\sigma} = 12.75$ ms. Good results have been obtained using these settings.

Additionally, the following special cases are accounted for:

during stimulation: updating of \bar{x}_i and σ_i^2 is interrupted, otherwise events shortly following a stimulation cannot be detected, primarily due to increased σ_i^2 -values.



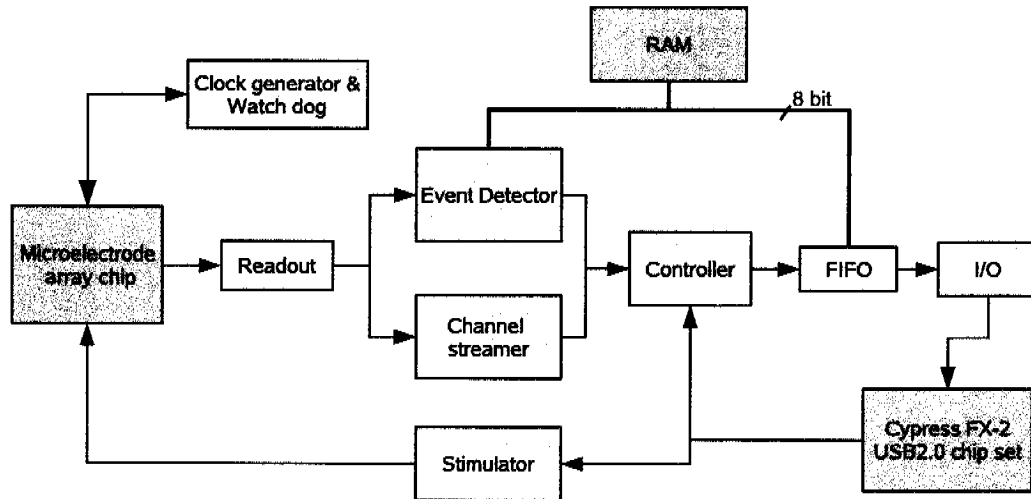


Figure 4.5: Simplified block diagram of the FPGA core. External components are the *microelectrode array chip*, then *RAM* and the *Cypress FX-2 chip set*. The *clock generator & watch dog* survey proper communication with the chip. Upon malfunction, the chip is reset, and the FPGA components perform a synchronous restart. There are three data paths: a first through the *event detector*, a second through the *channel streamer*, and a third through the *stimulator* in the reverse direction. The off-chip random-access memory is interfaced via a tri-state bus to two entities. Bus arbitration and address generation have been omitted for the sake of clarity.

$\sigma_i^2 \geq 64 (= 234 \mu V^2)$: no events are generated because the high level of noise indicates a defective channel.

$\bar{x}_i \geq 252 (= 480 \mu V) \vee \bar{x}_i < 4 (= -480 \mu V)$: no events are generated because the mean is too close to the rails and, therefore, no reliable calculation of σ^2 is possible.

The voltage values in brackets are valid for a gain of 1,000 and an ADC bias of 1 V.

Register-Transfer-Level Implementation Figure 4.5 shows how the event detection is embedded into the FPGA core architecture. Due to memory and bandwidth restrictions, the event length is fixed to a window of 256 samples, which corresponds to $256/20 \text{ kHz} = 12.8 \text{ ms}$. 128 of such look-back windows are maintained in the form of circular buffers on the external memory amounting to 32 kB of memory which clearly exceeds the Xilinx Spartan II XC2S200 on-chip block random-access memory (RAM) of 7 kB. Since all circular buffers buffer_i , $i = 0 \dots 127$, are synchronously used, it is sufficient to maintain only one buffer pointer, ptr , of 8 bit and to calculate the specific 15-bit buffer pointer using the relation $\text{ptr}_i = \text{ptr} + i \times 256$. The pointer, ptr , is incremented after each completed frame, i.e., after 128 samples. In the rare case of an event length larger than 256 samples, the beginning of the event will be clipped.

The respective 128 tuples of (\bar{x}_i, σ_i^2) , where $\bar{x}_i \in [0, 255]$ and $\sigma_i^2 \in [0, 127]$, are encoded using 14-bit fix-point numbers, where there are 6 and 7 binary digits to the right of the decimal point, respectively, and are then stored in the distributed on-chip RAM of the FPGA.

As can be seen in Figure 4.6, event detection is performed on the fly; samples are processed in the order as delivered by the neurochip. In order to compensate for RAM latencies induced by I/O operations, a 32-value deep first-in first-out (FIFO) buffer is placed upstream of the event detection. Upon availability of a new sample, x_i , from channel i , the respective (\bar{x}_i, σ_i^2) tuple is retrieved and updated with the new sample, x_i , and the i -th sub-FSM state is retrieved from the memory, all within 2 clock cycles. Then the sub FSM is executed (1 clock cycle) and, potentially, an event is generated (~ 520 clock cycles, depending on RAM bus availability¹), the new sub-FSM state is stored (1 clock cycle), and the FSM then returns into the *idle* state and is ready for the next sample.

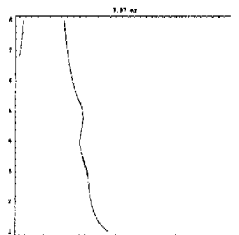
Benchmarking of the FPGA-design without the USB data transmission (which is not the limiting factor) in Modelsim [Mentor Graphics Corp., 2004] revealed a bandwidth sufficient to extract $50 \cdot 10^3$ events per second from 1024 channels. This spike rate corresponds to the order of magnitude that a neural burst on 1024 channels can reach during a network-wide burst. For 128 channels there is a margin of one order of magnitude. In case of a buffer overflow, events are dumped, and an error flag is activated and reported to the PC to indicate a loss of events. The major limiting factor is the bandwidth of the off-chip FPGA RAM, where the signal look-back windows are stored.

4.2.2.2 Stimulation

As described in section 4.2.1, the architecture of the chip allows for selecting an arbitrary set of electrodes for stimulation with the same stimulus waveform. Reconfiguration of the stimulation subset and reset of the filters proceed at the speed of the readout of the array, since the addressing scheme for readout is also used for programming of the electrodes: whenever an electrode is sampled for A/D conversion, either its “reset” state or its “stimulation” state may be reprogrammed. Consequently, the array may be reconfigured from the stimulation of a certain set of electrodes to the stimulation of any other set of electrodes within 50 μ s. In case that simultaneous stimulation of electrodes with different stimulus waveforms is desired, two consecutive simulations with a pause of 50 μ s need to be performed. When a stimulus of, e.g., 150 μ s length is applied, the two stimuli are 200 μ s apart, which, in most cases, can be considered to be simultaneous with respect to the neuronal culture.

Requirements for the stimulation control include:

¹The external RAM is used for the event detection and the USB 2.0 I/O data buffer. A bus arbitration unit and an address generation unit are implemented for the RAM bus assignment.



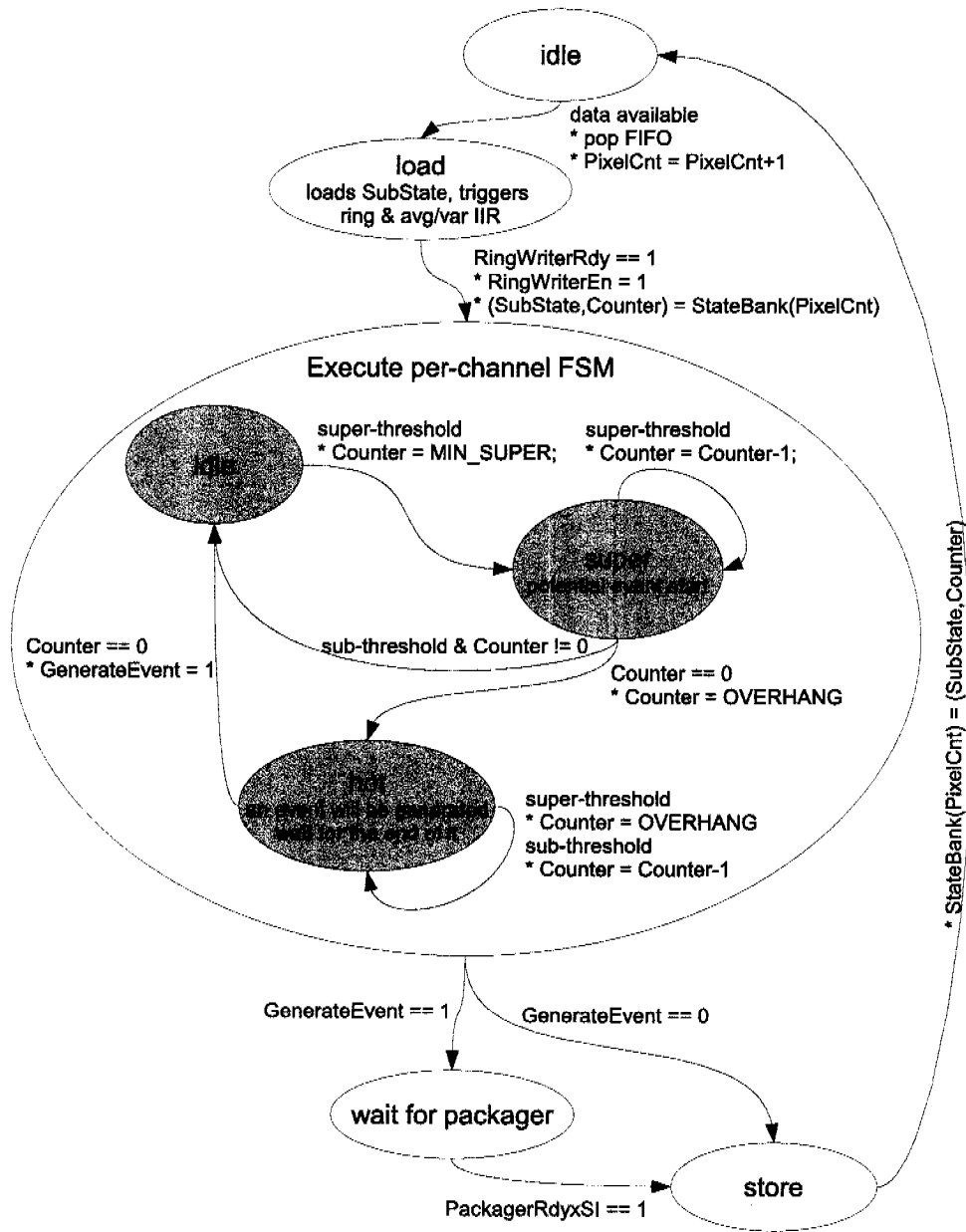


Figure 4.6: Finite state machine (FSM) as implemented on the FPGA to perform event detection. Annotations aside the state transition arrows have two meanings: If preceded by an asterisk, they represent an post condition, otherwise they indicate the premisc for the respective state transition. The FSM may be structured into two parts: an outer FSM and a sub FSM. The outer FSM manages the look-back ring memories, IIR filters for the floating average and variance calculation and retrieval of the per-channel sub-FSM state. The sub FSM performs the actual event detection per channel.

The sub-FSM state (3 bit, one-hot encoded) and the counter (5 bit) need to be stored per electrode. Whenever a new sample x_i arrives, the sub state is loaded into the sub FSM and executed into its next state.

- accurate timing: one frame or 50 μ s
- different stimulation electrode sets and various stimulus waveforms
- fast sequences of stimulations to facilitate time-coded stimulation.

Since the USB link latencies do not allow for a real-time issuing of commands, the FPGA needs to provide the infrastructure for accurate timing. A stimulation is specified by

- `start_frame`, relative or absolute (16 bit)
- `electrode_selection` (1 bit/electrode = 128 bits)
- `reset_selection` (1 bit/electrode = 128 bits)
- `waveform` (8 bit, 60 kHz, variable length).

A command providing the above-specified data is issued to the FPGA and enqueued in the FPGA's stimulation FIFO. The start frame triggers one of three possible timing modi *immediate*, *absolute*, or *relative*, the first two of which are self-explanatory, whereas *relative* means “stimulate $n \times 50 \mu$ s after the last stimulation.” The stimulation state machine walks through the following states:

Wait: wait for data. `start_frame` triggers one of three possible timing modi:

- `start_frame` = 0 \Rightarrow *immediate* stimulation
- $0 < \text{start_frame} < 2^{15} \Rightarrow$ *relative* stimulation: “continue `start_frame` frames after last stimulation”
- $2^{15} \leq \text{start_frame} \Rightarrow$ *absolute* stimulation: “continue when the lower 15 bits of the global frame counter (50- μ s clock) and `start_frame` match.”

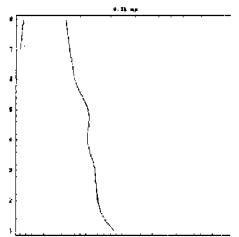
Program reset: at each electrode, program the filter reset switches according to `reset_selection`

Program stimulation buffers: at each electrode, program the stimulation buffers' on/off and stimulation switch according to `electrode_selection`

Replay waveform: stimulate according to `waveform`

Program stimulation buffers: switch all stimulation buffers off and disconnect them from electrodes

Program reset: switch all reset switches off, back to *Wait*



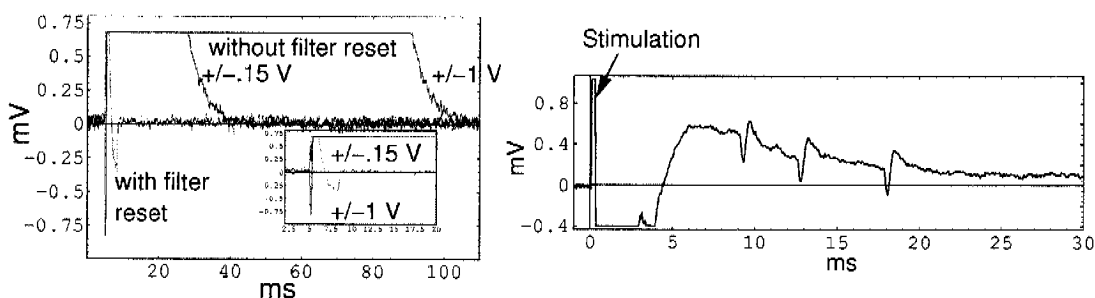


Figure 4.7: Left: Efficiency of the on-chip reset function for stimulation. Bipolar pulses (220 μ s overall duration) of different amplitudes (± 0.15 V and ± 1 V) in saline solution have been applied, and the reset has been operational or not. Without reset, it takes 20 to 100 ms for the recording circuitry to return to the measurement range. The inset shows a close-up of the initial 20 ms at better temporal resolution.

Right: Stimulated signals from rat cortical neurons on the stimulation electrode. The stimulus is a biphasic rectangular pulse of 1.1 V_{pp} and 110 μ s duration. The remaining slow drift can readily be eliminated by digital high-pass filtering. Both measurements have been performed with 20- μ m-diameter platinum electrodes.

Stimulation artifact suppression Figure 4.7 shows the efficiency of the on-chip reset function for stimulation artifact suppression. In Figure 4.7, left, bipolar pulses (220 μ s overall duration) of different amplitudes (± 0.15 V and ± 1 V) in saline solution have been applied, while the reset has been operational or not. The recordings from the stimulation electrode are shown. The readout filter for the red and green trace was reset 50 μ s before the stimulation pulse was applied and kept until 50 μ s after the stimulation pulse was finished. Without reset, it takes 20 to 100 ms (depending on the stimulation amplitude) for the recording circuitry to return to the measurement range. The inset shows a close-up of the initial 20 ms at better temporal resolution. With reset, it takes less than 5 ms for the recording circuitry to return to the measurement range. It is worth mentioning that a stimulation sequence always ends with a stimulation of a value close to the equilibrium potential of the electrode so that the electrode quickly returns to its equilibrium potential after stimulation.

A similar experiment on a chip with rat cortical neurons is documented in Figure 4.7, right. The stimulus is a biphasic rectangular pulse of 1.1 V_{pp} and 110 μ s duration. The measurement shows action potentials recorded on the stimulation electrode 9 ms after stimulation. Action potentials have been recorded less than 5 ms after stimulations on the same electrode. The remaining artifact signal shows very slow temporal characteristics and can readily be removed by digital highpass filtering. The other electrodes that do not provide a stimulation signal record the stimulation signal as a spike: the signal is sufficiently attenuated in the culture bath so that the signal is too weak and too short to impact the respective highpass filters.

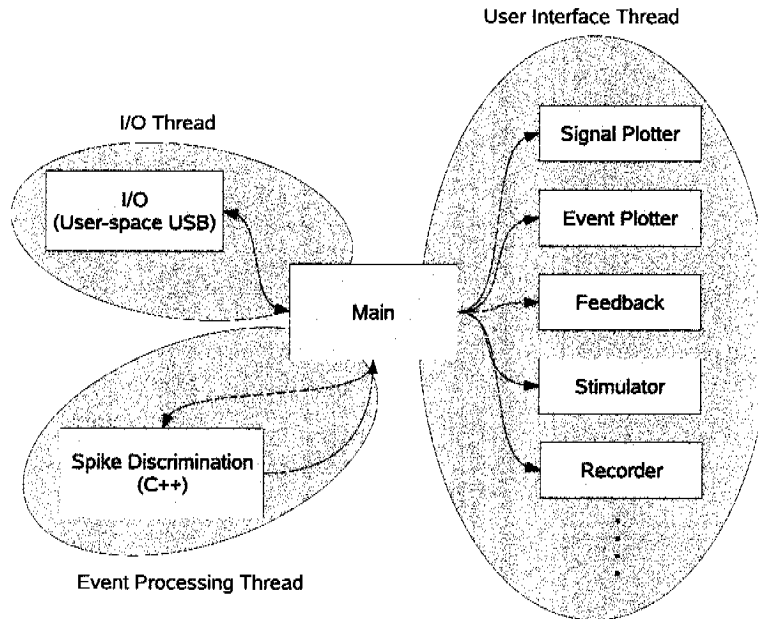


Figure 4.8: Software modules and threads. The *main* module contains semaphores for managing the data exchange between the threads (ovals). Each module (boxes) except for the *main* module is interchangeable during runtime. The USB-I/O module can be, e.g., exchanged for a *replay* module that will replay previously recorded data. Apart from the three main threads, further threads may be generated inside the modules.

4.2.3 Software

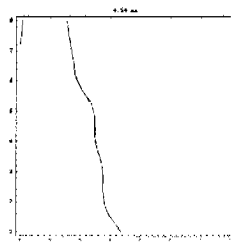
The software is a crucial component of the system. It organizes the USB data handling, the spike sorting, the data visualization, the stimulus generation and the automated feedback generation. Low latency time and high data throughput are key requirements. It is implemented in C++ using the Qt library [Trolltech, Inc., 2007] for the graphical user interface and runs on Linux.

There are three concurrent, widely independent tasks: the I/O-handling, the event processing, and the user interface. Therefore, the program named *Neurotalker* runs in 3 main threads as sketched in Figure 4.8.

4.2.3.1 I/O

The I/O thread interfaces the USB subsystem to the main loops of the program. Since USB is a pure Master-Slave bus USB Implementers Forum [2005], and since it relies on block transfers, a further protocol layer is required to have the capability to transfer data packages of arbitrary size at any defined time.

A combined software and VHDL package, named *ce-usb-kit* that provides this additional protocol layer has been developed and encapsulated into an independent library containing both VHDL and C code. It is available from the *sourceforge* project Hafizovic [2005]. The *ce-usb-kit* provides a VHDL code that includes



- an external-RAM I/O buffer,
- a receiver finite-state machine that decodes data packages sent by the PC,
- a sender finite-state machine that encodes data, performs padding and calculates an 8-bit cyclic redundancy checksum (CRC).

The 8-bit CRC is appended to each transferred data block and is very useful for testing purposes, especially with respect to the communication between FPGA and FX-2 chip. The software part of the *ce-usb-kit* runs exclusively in the user-space and provides routines to handle the USB device. The communication between PC and FX-2 relies on exchanging data with the FX-2 firmware as provided by CESYS via I/O-controls. Data are sent in packages of multiples of 512 bytes, i.e., after a bulk transfer request for 4096 bytes is issued, the FX-2 has to deliver the requested amount of data. In the case that less data is available than requested, the excess volume is padded. The polling interval and the package size are dynamically regulated to ensure an efficient data transport. The output buffer on the FPGA is dimensioned to 32 kB.

4.2.3.2 Modules

A number of universal modules provide the infrastructure to perform standard tasks. The list of modules includes

Signal plotter: plots data of arbitrarily selectable channels

Rasterplot: plots events in a rasterplot

Stimulator: stimulates via an arbitrary set of electrodes

Recorder: stores the channel data and the events into files

Feedback: triggers actions, i.e., issues stimulation signals or a beep upon events matching certain criteria

LUT programmer: programs the Look-up table (LUT) that controls the multiplexing of the A/D converters to the electrodes. Sampling rates of up to 160 kHz are possible, when only one electrode per row is read out.

I/O modules include the *USB-I/O* and the *replayer*. The *replayer* streams in data that have been previously recorded with the *recorder* module.

It is possible to seamlessly add further modules owing software design, which includes techniques like instantiation by name [Taschini et al., 1999]. Dedicated modules for special experiments can be generated and added.

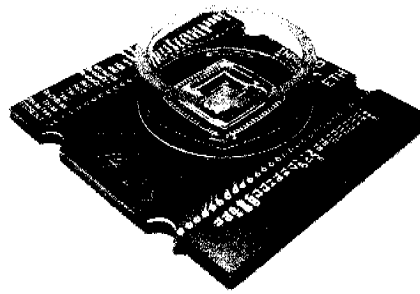


Figure 4.9: Photograph of the packaged chip. The chip has been glued to a PCB and wire-bonded. The glass ring has been mounted on top, and the bond wires have been covered with a biocompatible epoxy resin, EPOTEK 302-3M [Epotek].

4.2.4 Fabrication

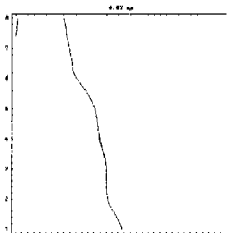
The chip has been fabricated in industrial 0.6- μm CMOS-technology [XFAB]. The platinum electrodes and the additional passivation layers have been fabricated in house using a 2-mask postprocessing as described in Heer et al. [2004]. Thereby, a variety of CMOS MEAs differing in electrode (Figure 4.2) diameters (10 to 40 μm) and locations (pitch 50 to 500 μm) could be fabricated on the same CMOS chip.

The chip operating conditions are the same as for conventional cell incubation, i.e., 95% rel. humidity, saline solution as cultivation medium and a temperature of 37°C over extended time (successfully tested during 9 months), so that a biocompatible package and inert electrical contacts are required. The array chip has been mounted and wire-bonded to either a ceramic package or onto a PCB, both featuring gold contacts. Afterwards, a glass ring has been mounted, and the bond wires have been encapsulated with epoxy resin (EPOTEK 302-3M [Epotek]). Figure 4.9 shows a ready-to-use packaged chip. Additionally, the electrodes can be electroplated with porous platinum black to reduce the electrode impedance [Heer et al., 2004].

The workflow for the FPGA core generation encompasses the VHDL code entry Ashenden [2002], the simulation with Modelsim Mentor Graphics Corp. [2004], the synthesis with Synplify Synplicity [2005], and, finally, the layout with Xilinx ISE Xilinx [2005].

4.3 Results

Electrical testing of the on-chip bandpass filter revealed a tuning range for the lower corner frequency between 1 Hz and 1 kHz and for the upper corner frequency between 1 kHz and 30 kHz. The midband gain of the bandpass was measured to be 19.12 dB. The total-equivalent-input noise was found to be 5.9



μV_{rms} (10 Hz - 100 kHz) at a gain of 1,000. This is comparable to the thermal noise of the platinum electrodes in physiological saline solution which was measured to be $3.19 \mu V_{\text{rms}}$ for $10 \times 10 \mu\text{m}^2$ and $2.16 \mu V_{\text{rms}}$ for $40 \times 40 \mu\text{m}^2$ electrodes in a band between 10 Hz and 100 kHz. Thus the total noise of the setup is dominated by circuitry noise, which is somewhat higher than the thermal noise.

The measured power consumption was 120 mW at 5 V supply, 20 mW of which are dissipated within the electrode array. The increase of the temperature within the electrode array was measured to be less than 1°C so that cooling is not required.

4.3.1 Stimulated Activity in Rat Cortical-Neuron Cultures

As a demonstration of precise spatio-temporal stimulation patterns and event detection, we present the following experiment inspired by the liquid state machine concept as presented in [Maass et al., 2002, 2003]. The underlying idea is to use the stimulation responses of arbitrary excitable media for information processing. As shown in Figure 4.10 the model consists of a stimulator, a neuronal network on a neurochip, an event detector, a leaky integrator and a classifier.

4.3.1.1 Method

Stimulation Two stimulation spots a and b have been defined, where a represents a stimulation on electrode 52 and b stimulates on electrodes 96-104, both with one biphasic rectangular pulse of $1.1 V_{\text{pp}}$ and $110 \mu\text{s}$ duration. In Figure 4.2 location a corresponds to the fifth electrode in the third row from the bottom and location b to the sixth row of electrodes from the top. For this experiment, the stimulation pulse amplitude of $1.1 V_{\text{pp}}$ was found to be the lowest voltage that reliably evoked action potentials.

Based on these two stimuli, the two spatio-temporal patterns aba and bab have been stimulated with an inter-stimulus period of 40 ms. The inter-stimulus period of 40 ms was chosen to be on the order of the neuronal time constant, which is in the range of 20 ms to 100 ms [Kandel et al., 2000, p. 222], to allow for temporal summation to occur. The period in between aba and bab stimulations, the inter-pattern period, was chosen to be 2.3 s so that it is short enough to suppress spontaneous activity. Without stimulation interference, spontaneous activity occurred at intervals of 3 to 5 s.

As described in section 4.2.2.2 on page 33, the stimulation patterns are generated on two levels: the inter-stimulus period of 40 ms is generated on the FPGA with $50 \mu\text{s}$ accuracy, and the inter-pattern period of 2.3 s is generated on the PC with millisecond accuracy.

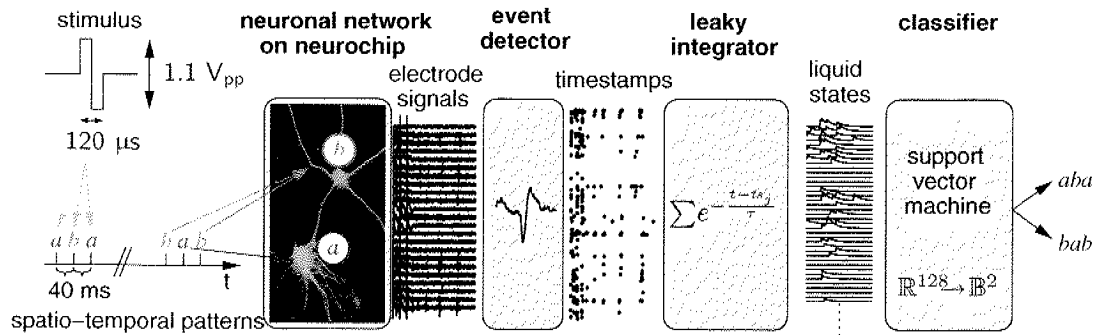


Figure 4.10: Schematic of a “liquid state machine” implementation. Inputs, either *aba* or *bab*, are stimulated in the neuronal culture. From the neuronal culture response, liquid states are calculated by leaky integration of the timestamps. 200 ms after stimulation, the liquid state is sampled and classified in a support vector machine.

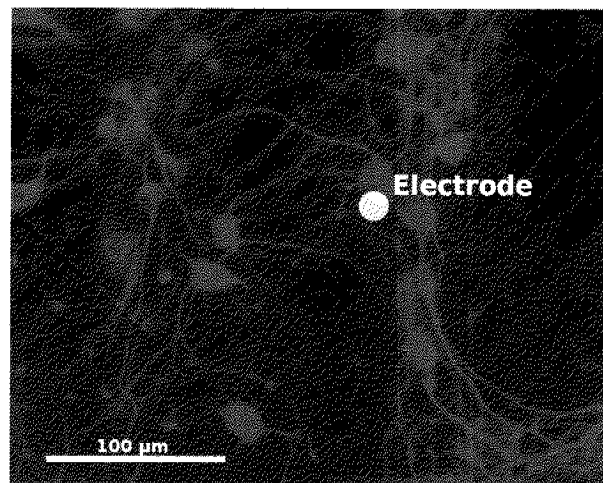


Figure 4.11: Rat cortical neurons on the CMOS MEA after 17 days *in vitro*. Neurons are rendered visible (green) by immunostaining of the microtubuli-associated protein-2 (MAP2) with a fluorescein-isothiocyanate (FITC)-conjugated secondary antibody.

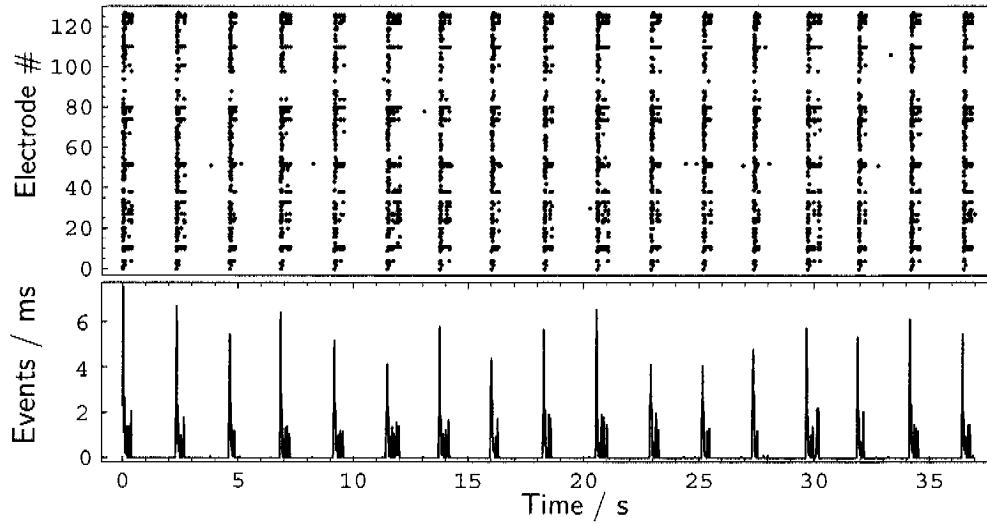


Figure 4.12: Rasterplot and network-wide firing rate of the *aba* stimulation pattern repeated at 2.3 s.

With regard to potential damage of the stimulated neurons, we noticed that after more than 200 stimulations on the same electrode, the cells could still be stimulated without any observable response signal deterioration.

Neuronal Culture Hippocampal tissue was extracted from newborn Sprague-Dawley rats and dissociated by trypsinization (0.25% Trypsin / EDTA) and gentle trituration. Cells were plated on laminin (10 $\mu\text{g}/\text{ml}$) and Poly-L-Lysine (10 $\mu\text{g}/\text{ml}$) coated CMOS chips at a density of 15,000 cells/ mm^2 and held in Neurobasal medium (Gibco) containing B27 supplement (2%, Gibco), fatty acid supplement (0.1%, Sigma), lipid mixture (0.1%, Sigma), Alanyl-Glutamine (2 mM, Gibco) and sodium pyruvate (1 mM, Sigma) (Figure 4.11). Measurements were undertaken after 37 days in vitro. For all biological measurements, the packaged chip, as shown in Figure 4.9, (20- μm -diameter electrodes, pitch 250 μm) was mounted onto a heated (37°C, manually adjusted heating power) PCB in ambient atmosphere, as shown in Figure 4.1.

Readout From the detected events (described in section 4.2.2.1), 128 sets, ts_i , of timestamps have been generated, where $i = 0 \dots 127$ designates the channel. From each of these sets, ts_i , a liquid state function

$$x_i(t) = \sum_{ts_i: t_{\text{cutoff}} < t - ts_i \leq 0} e^{-\frac{t - ts_i}{\tau}} \quad (4.3)$$

has been synthesized, where $\tau = 20 \text{ ms}$ is the leak time constant and t_{cutoff} is 40 ms. t_{cutoff} explicitly limits the “memory” of the readout function. For each

stimulation experiment, j , a time snapshot of the liquid state,

$$\mathbf{x}_j = \mathbf{x}(t_{\text{stimulus},j} + 200 \text{ ms}),$$

has been calculated. These 128-dimensional response vectors represent the neuronal network response to the stimulation pattern that was initiated 200 ms before.

A direct information transfer from the stimulator to the readout that might bypass the neuronal network is excluded, by choosing the \mathbf{x}_j -sample time as $t_{\text{stimulus},j} + 200 \text{ ms}$, with $t_{\text{cutoff}} = 40 \text{ ms}$, and having the last stimulation at $t_{\text{stimulus},j} + 80 \text{ ms}$. With these settings, there is a time span of $200 \text{ ms} - t_{\text{cutoff}} - 80 \text{ ms} = 80 \text{ ms}$, where the neuronal network is the exclusive mediator of any information. Thereby, the function of the neuronal network in the liquid state machine is defined.

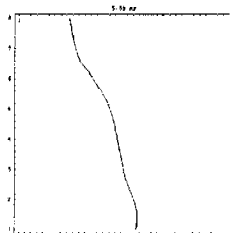
Classification of the responses, \mathbf{x}_j , with respect to the preceding stimulation pattern *aba* or *bab* has been performed off-line using Mathematica Wolfram Research [2005] and a support vector machine package [Nilson et al., 2006]. For the presented experiment, a total of 44 stimulation-response pairs, 18 *aba* stimulations succeeded by 26 *bab* stimulations, have been recorded. From each of the two classes the first 30% (6 and 9) of the pairs was used to generate a weight vector, $\mathbf{w} \in \mathbb{R}^{128}$, and a *bias* $\in \mathbb{R}$. With the linear kernel used, \mathbf{w} can be seen as the normal vector of a 128-dimensional hyperplane given by $\mathbf{x} \in \mathbb{R}^{128} : \mathbf{w}^T \cdot \mathbf{x} + \text{bias} = 0$, where T denotes transposition, that separates the *aba* and *bab* responses. Hence, the sign $(\mathbf{w}^T \cdot \mathbf{x}_j + \text{bias})$ represents the classification of \mathbf{x}_j .

4.3.1.2 Results

As can be seen in Figure 4.14, the two classes of the respective first 30% of stimulation-response pairs that have been used as training set are *separable*. The remaining two thirds of each of the two classes, in total 29 responses, constitute the test set and have been classified according to the weight vector \mathbf{w} and *bias* as obtained from the training with the first 30% stimulation-response pairs. From the test set, 83% (24 of 29) were correctly classified. 75% (9 of 12) of the *aba* test set and 88% (15 of 17) of the *bab* test set have been correctly classified. Figure 4.14 also shows the results graphically.

4.3.1.3 Discussion

This first experiments indicate that information processing with dissociated neuronal networks according to the liquid state machine model might be possible. There are still many issues to be addressed, e.g., the sequence of the *aba* and *bab* patterns should be random so that the effects of other transient changes in the neuronal culture can be eliminated. Also, the measurement time needs to be



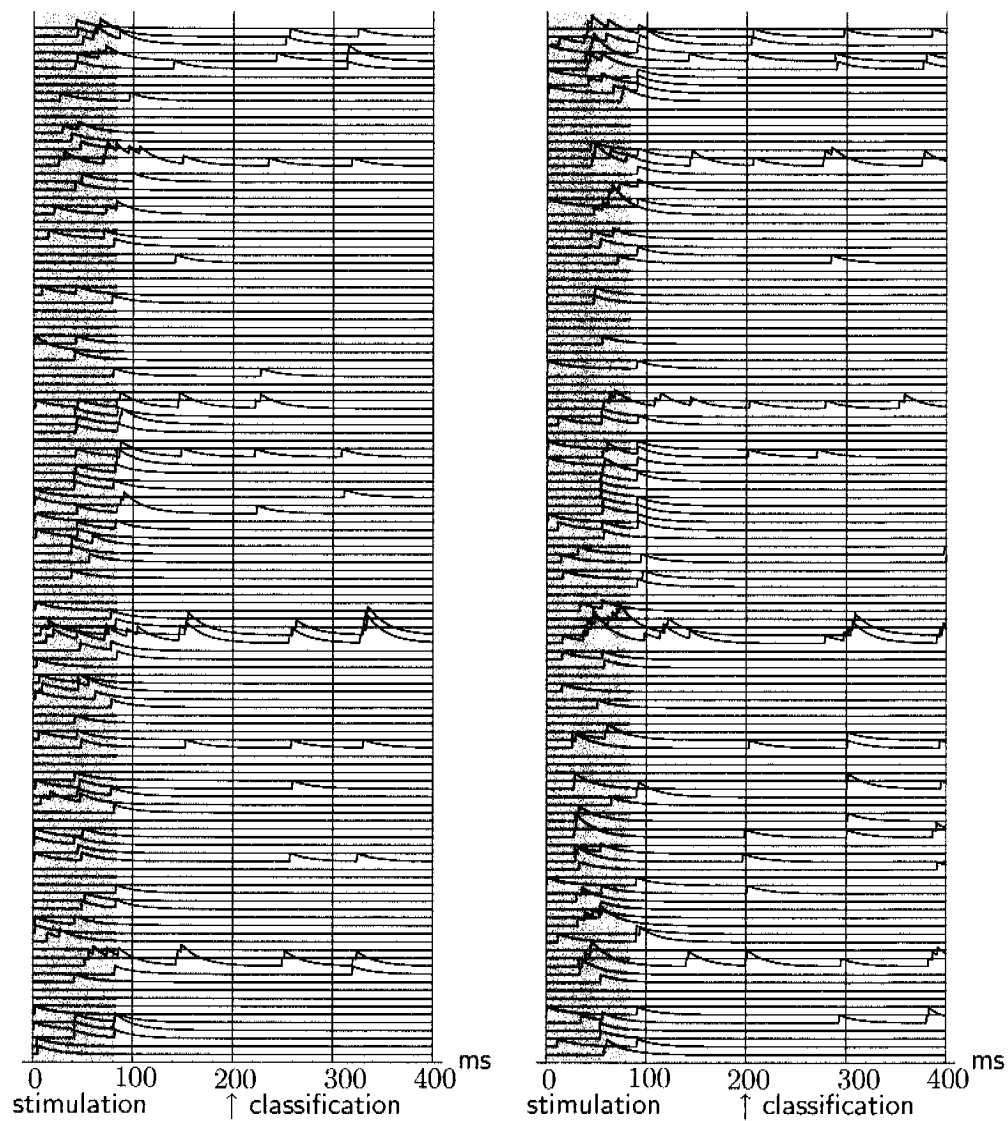


Figure 4.13: Plot of the outputs of the leaky integrators, $x_i(t)$, also referred to as liquid states, aligned to a stimulation. The left plot shows an *aba* stimulation, the right plot shows a *bab* stimulation. The gray background indicates the time span during which the three stimulations on the a and b sites at 0 ms, 40 ms and 80 ms take place. 200 ms after the first stimulation, the 128 leaky integrators are sampled, and the obtained vector is then classified.

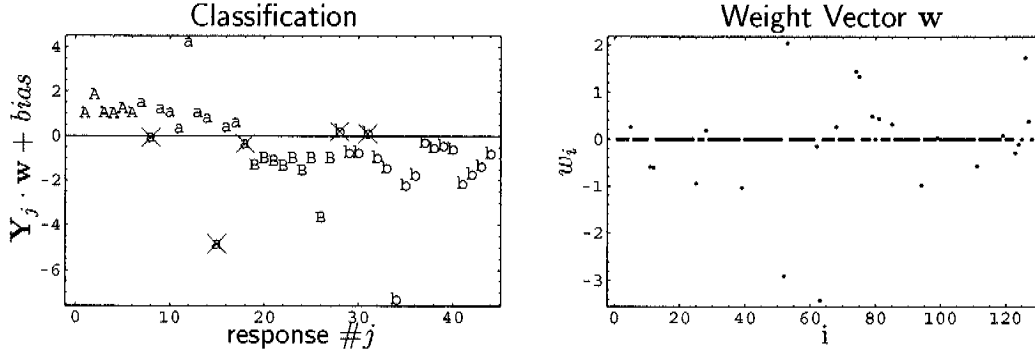


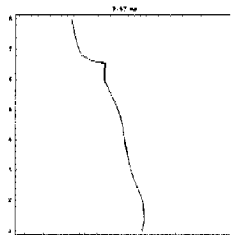
Figure 4.14: Classification of neuronal responses. Left: Plot of the $\mathbf{w}^T \cdot \mathbf{x}_j + \text{bias}$ values, the sign of which denotes the classification. *aba* responses denoted by “a” are expected above the zero line and *bab* responses denoted by “b” are expected below 0. The capital letters, A, B, denote the training data. Crossed-out letters denote false classifications. All training data are correctly classified, i.e., the training data is *separable*. 83% of the test responses (i.e., $1 - \frac{\#\text{“crossed-out small capitals”}}{\#\text{“small capitals”}}$) have been correctly classified.

Right: Plot of the components w_i of the weight vector \mathbf{w} . Only about 20% of the w_i are non-zero. Therefore, only about 20% of the electrode signals actually contribute to the classification.

extended by performing the measurements in an incubator to be able to better characterize each of the stimulation sites *a* and *b*. These issues and other possible shortcomings will be addressed in future work.

4.4 Conclusion

The presented signal processing chain reaches from the cell culture, which lives on top of a CMOS-based microelectrode chip to the C++ control and analysis software. The latter enables a bidirectional communication with electrogenic cell cultures *in vitro*. The combination of the CMOS chip with an FPGA-based device allows to perform online event detection and, on demand, full-depth transmission of all 128 20-kHz signals. The event detection is implemented on the FPGA running at 48 MHz FPGA-clock speed, is fully self-adjusting and requires no user inputs. The data volume transferred to the PC is effectively decimated by relocating the event detection from the PC to the FPGA. The system is capable of identifying an event and triggering a stimulation within less than 2 ms after the event has been detected, so that it may be considered real-time with respect to neuronal networks cultured on the chip.



4.5 Acknowledgements

The authors thank Prof. Henry Baltes for sharing laboratory resources and for his ongoing stimulating interest in their work. The authors are indebted to Frauke Greve, ETH Zurich, for help and advice on neuronal cell cultures, Beat Hofstetter for his contribution to the Neurotalker software and Urs Frey, ETH Zurich, for many fruitful discussions. Prof. Wolfgang Maass, university of Graz, Austria, is acknowledged for stimulating discussion on liquid state machines and liquid state classification. Funding has been generously provided by the European “Information Society Technologies” “Future and Emerging Technologies” program, and the Swiss Bundesamt für Bildung und Wissenschaft (BBW), contract number IST-2000-26463.

Chapter 5

Information Processing with Natural Neuronal Networks

Many of the current activities in applied neuroscience can be categorized in

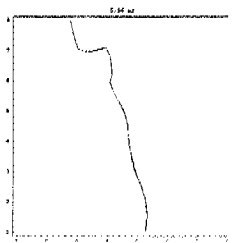
1. developmental / anatomical,
2. single-neuron-level, and
3. culture-level

research. Developmental research has great importance especially in view of applications in medicine. The single-neuron-based research is the classical, (cell-) biological approach and is key to understand how regulatory mechanisms and synaptic plasticity function on the cell and regulatory level. Culture-level investigations, where the emerging collective properties are of primary interest, are prominent also in informatics and physics.

The work presented in this thesis clearly belongs to the last group of culture-level research. With the developed neurochip, a more sophisticated tool to investigate the complex, emergent behavior of *in vitro* neuronal cultures and of brain slices is provided.

5.1 Theoretical Neuroscience

Theoretical neuroscience spans over several levels of abstraction that are comparable to those in applied neurosciences: it reaches from systembiological modeling [Mendes, 1997], over single-cell simulations [Kerszberg and Changeux, 1998, Hines and Carnevale, 1997], to artificial networks. Research on *in vitro* natural neuronal networks (NNN) can significantly benefit from research on artificial



networks. An example for such synergies is given in this thesis by the application of the liquid state machine concept to NNN. Presumably, there will also be transfers in the other direction: a model that works for NNNs is likely to be also applicable to an artificial network.

It is an interesting observation that as artificial networks evolved in the direction of biological systems, and concepts like noise have been introduced, the methods used to train and describe these artificial networks became more applicable to natural neuronal networks. Back-propagation for example is still a very effective method to train deterministic artificial networks but is neither applicable to non-linear artificial networks nor does it produce much insight into biological systems. With the development of increasingly non linear artificial neuronal networks, where a Jacobian is not available, training concepts with less premises are needed. Weight perturbation is one of these model-free training methods [Dembo and Kailath, 1990]. Other approaches apply a hybrid strategy by reducing the number of model assumptions, see [Fiete and Seung, 2006].

5.2 Computing without States - Liquid State Machines

A prominent concept for computing with unstructured neural networks that provides a mathematical framework is the Liquid State Machine (LSM) presented by Maass, Natschl ger, and Markram [2002], a review-style description is given by Florian [2003]. It describes a concept how to process information without digital finite states.

The LSM is composed of a nonlinear filter, L , (a function that maps time-varying input streams $u(\cdot)$ to other time-varying streams $x(\cdot)$), and a potentially memoryless readout function, f , that maps the “liquid” state $x(t)$ of the filter to a target output $y(t)$. The filter, L , has a fading memory: $x(t)$ may depend not only on $u(t)$, but also, in a quite arbitrary, nonlinear fashion on previous inputs $u(t - \tau)$. Mathematically speaking, $x(t) = (Lu)(t)$ and $y(t) = f(x(t))$ for the memory-less readout function or $y(t) = f(x(t - \tau))$, $\tau \in [0, t_{\text{mem}}]$ for a readout function with a memory limited by t_{mem} .

The LSM features universal computational power for time series, if two simple conditions are met.

Separation Property. The filter, L , needs to be capable to separate two different inputs, $u_1(t)$, and $u_2(t)$ into two different trajectories and to produce two different outputs $x_1(t)$ and $x_2(t)$ (the pointwise separation property). The separation property closely related to the deterministic and chaotic properties of the filter L . (There exists a number of metrics to quantify the separation property, e.g., the Ljapunov exponent and mobility of information [Chirikov, 1979, Ramasubramanian and Sriram, 2000].)

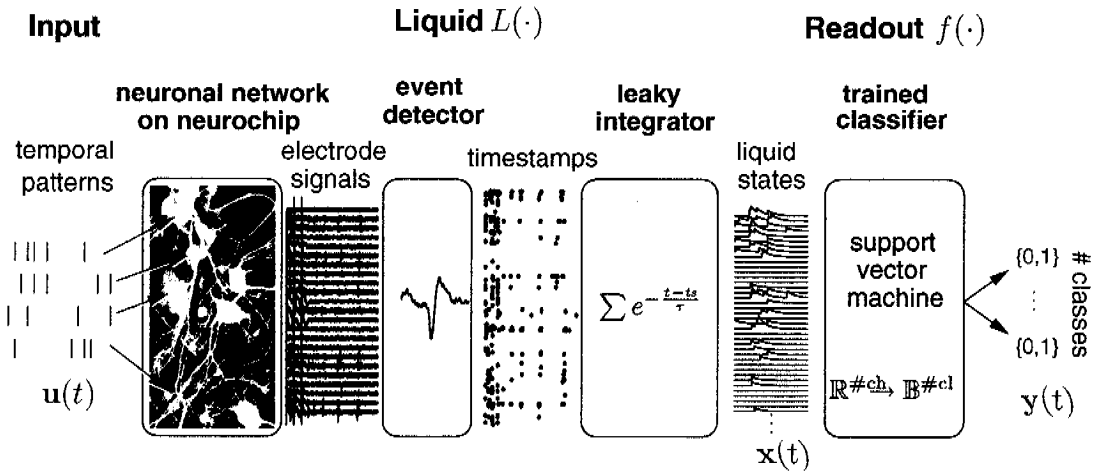


Figure 5.1: Structure of a liquid state machine (LSM). The numbers of channels and classes are denoted by $\#ch$ and $\#cl$.

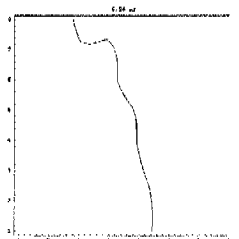
Approximation Property. The ability of the read-out device to probe the computation media’s state accurately enough to achieve a sufficient approximation of it’s internal state.

5.2.1 Computing with Unstructured Networks?

Results with artificial spiking networks have shown that a randomly connected network of inhibitory and excitatory synapses is capable of efficient information processing. An impressive demonstration is given in [Maass et al., 2003], where Maass et al. tackle the speech recognition challenge posed by [Hopfield and Brody, 2000]. Hopfield’s and Brody’s challenge consisted of recognizing the 10 words “zero”, “one”, ... “nine”, each spoken 10 times by 5 different speakers, with each spoken word encoded into 40 spike trains. The result was that the performance of an unstructured, randomly generated network of 135 integrate-and-fire neurons embedded into an LSM was similar to that of the ingenious circuit consisting of more than 6,000 integrate-and-fire neurons constructed especially for this task by Hopfield and Brody.

Even though often referred to as state-less computing, information transmission by action potentials (neglecting the unreliable synaptic transmission) still features the digital characteristics of all thermodynamically stable and noise robust systems. But, unlike information stored in, e.g., DNA or synapses, the current state of activity of a neuronal network is not stable over time but has a fading character.

Promising applications can be found in the classification of temporal signals on the second time-scale, such as voice recognition.



5.2.2 Prior Art of LSM

The prior art is largely confined to computational experiments. The original LSM [Maass et al., 2002] used recurrent spiking neural networks as filters L . Other recent LSM implementations include a finite element simulated model of a liquid [Goldenholz, 2003], or even real water in a bucket [Fernando and Sojakka, 2003].

5.3 The Liquid State Machine Concept for Neurochips

The most fundamental differences with respect to the simulation experiments investigated by Maass et al. [2002], arise from the biological nature of the neuronal network. Neurons migrate, synaptic changes occur, and environmental changes influence the NNN function: the computation medium is not static.

In contrast to simulation experiments, where the readout is taken from a single defined neuron, the electrode-neuron contact is much less reliable. At a first glance, a neuron-electrode interface may very much look like a synapse: action potentials are passed from a neuron to the electrode or in the other direction. However, there is no synaptic character in the way that there is no means to, e.g., influence the “synaptic plasticity”: the electrode cannot attract a particular neuron.

For this reason it is necessary that the readout mechanism is constantly re-adjusted in an automatic fashion. The p-delta learning rule [Duda and Hart, 2001, Auer et al., 2002], which was used in [Maass et al., 2003] features these desired capabilities: It defines two thresholds: a *classification threshold* and a lower *correction threshold*. Whenever a response could be classified but was found outside the *correction threshold*, the readout weights have been tuned. However, it was found not to be very reliable mainly for two reasons: it tends to “forget” patterns as it is taught new ones and it shows the tendency to overfitting. Especially the first reason is problematic for a real-world use: upon prolonged exposure to the same pattern, it would improve its recognition, but at the same time, would forget the precedent training patterns. The overfitting problem is somewhat related: as a consequence of (over-) optimizing the system, it becomes more vulnerable to fluctuations in the NNN.

5.4 Results

In this section, further results of the experiment reported on in section 4.3 on page 39 are given. In section 4.3, the liquid state, $\mathbf{x}(t)$, was always sampled

exactly 200 ms following the stimulation, and the thus attained sample was classified with respect to the stimulation pattern. This may prove that information processing with NNN is generally feasible, but a synchronization of the readout to the input is not always possible in a real-world application. The LSM might, e.g., receive input from a microphone, where the time of the beginning of the stimulation is not available. Therefore the LSM should be able to determine itself, when a word has been spoken.

Maass et al. [2003] coined the term “anytime computing” to describe such a synchronization-free approach. In the following, the need for synchronization of the readout is abandoned, and a method that continuously samples and classifies the liquid state is introduced. First, in section 5.4.1, the temporal characteristics of the stimulation responses are investigated and then, in section 5.4.2, anytime computing is introduced. Again, leaky integrators are used to generate the liquid states $\mathbf{x}(t)$, like in section 4.3, and a support vector machine is used to classify those liquid states.

5.4.1 Liquid State Classification with Many Samples

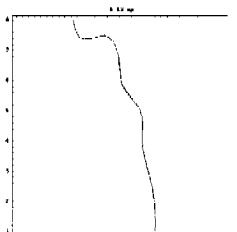
The capability to classify liquid states not only at a single given point in time, but over a time span from 160 ms to 300 ms after stimulation is discussed in this section. The same data as in section 4.3 have been used. For classification, the same support vector machine

$$c(j, t) = \mathbf{w}^T \cdot \mathbf{x}(t_{\text{stim},j} + t) + \text{bias} \quad (5.1)$$

is used, where $\text{sign}(c(j, t))$ is the decision rule.

The training set consists of $\mathbf{x}(t_{j,i})$ with $t_{j,i} = t_{\text{stim},j} + 160 \text{ ms} + i \times 10 \text{ ms}$ and $j = 1 \dots 6, 19 \dots 28, i = 0 \dots 14$, i.e., the time from 160 ms to 300 ms after stimulation is sampled every 10 ms. From the total of 44 stimulation-response pairs (18 *aba* stimulations succeeded by 26 *bab* stimulations) the first 30% (6 and 9) of the pairs were used for training, like in section 4.3. However, the number of training samples, $\mathbf{x}_{j,i}$, has been increased from 15 to 225 due to the multiple sampling of each stimulation pattern response. In contrast to section 4.3, the training set is not separable: no hyperplane exists that would separate all of the *aba* and *bab* pattern responses. This can also be seen later in Figure 5.4, where it is shown that misclassifications even of training data occurs. Furthermore, it is noteworthy that the weight vector, \mathbf{w} , shown in Figure 5.2, has considerably changed in comparison to Figure 4.14 on page 45: the weight vector is less sparse than before, which means that more readout channels (electrodes) contribute to the classification.

Figure 5.3 shows the resulting classifications at each sampling time. There is no defined point in time after stimulation that seems to be preferred for classification. Figure 5.4 gives an overview of the classification performance per



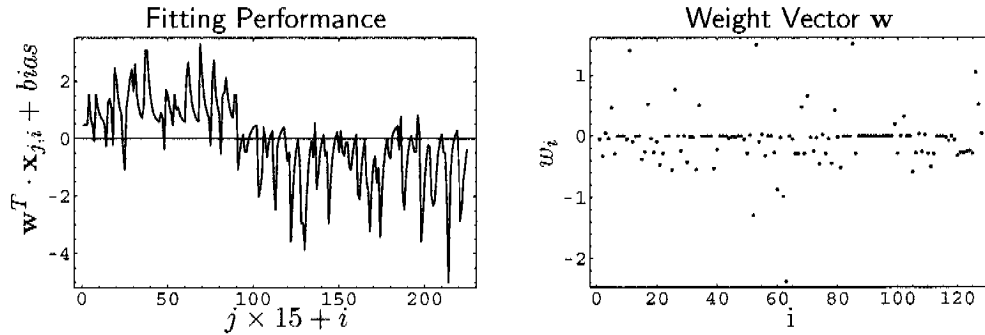


Figure 5.2: Left: Classification of the training set. The first 90 data points are expected above zero, the remaining data points below zero. There are misclassifications that are clearly visible so that the training set is not separable. Right: The weight vector, \mathbf{w} , is more dense than the weight vector used for the classification in section 4.3 (Figure 4.14 on page 45). More readout channels contribute to the classification.

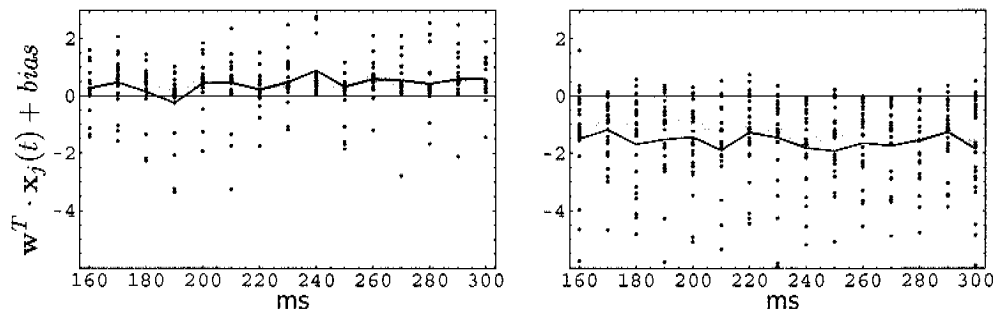


Figure 5.3: The left and right plot show the classifications of the NNN responses to the *aba* and *bab* stimulation patterns. Additionally, the median, grey line, and the mean, black line, of all responses at the given stimulation-referred time, are plotted. There does not seem to be any defined time after stimulation that provides better classification performance in this configuration.

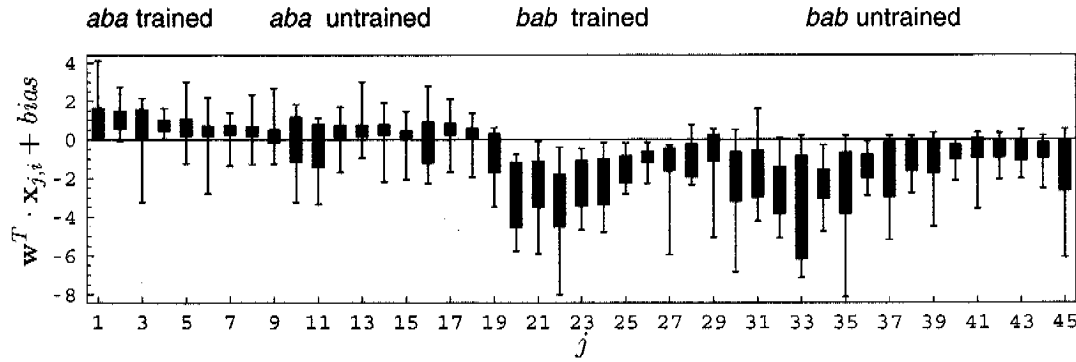


Figure 5.4: "Box-whisker" plot of each stimulation pattern classification. The Box whisker plot takes the form of a box that spans the distance between the 25% and 75% quantiles. The inner line denotes the median and the "whiskers" extend to span the full data set. The median corresponds to a majority vote: if the majority of the classification is above zero, it will also be above zero. For *aba* and *bab* stimulations, the median is expected above and below zero.

stimulation experiment j . It shows that there are big differences, especially in the classification spread within each stimulation response, j , over the stimulation experiments.

Since there are 15 classifications per stimulation response (Figure 5.3), a method to yield a final classification based upon these 15 classifications is needed. There are many methods, among which are

1. leaky integration with thresholding,
2. integrate and fire, and
3. low-pass filtering with thresholding.

They all have in common that they consist of some variations of averaging and thresholding.

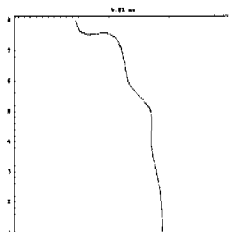
5.4.1.1 Mean-based classification

One approach is to take the mean value of the classifier and to decide based on its sign. Starting 160 ms after each stimulation, the value of the classifier is added to an accumulator:

$$f_{\text{mean},j}(t) = \sum_{i=0}^{(160+i \times 10)\text{ms} \leq t} c(j, (160 + i \times 10) \text{ ms}) \quad (5.2)$$

$$y_{\text{mean},j} = \text{sign}(f_{\text{mean},j}(300 \text{ ms})) \quad (5.3)$$

Figure 5.5 shows the resulting trajectories of the accumulator versus time.



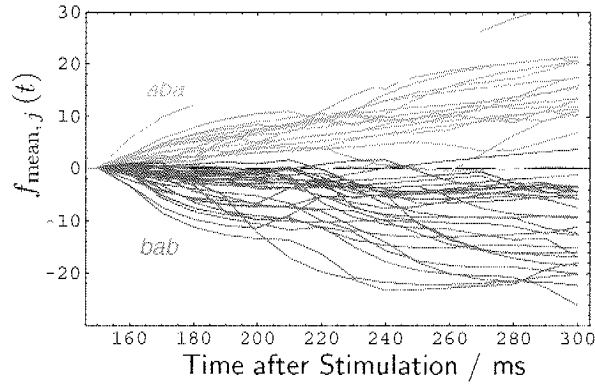


Figure 5.5: Mean-based classification. Plotted are $f_{\text{mean},j}(t) = \sum_{i=0}^{i \times 10 \text{ ms} \leq t} \mathbf{w}_i^T \cdot \mathbf{x}(t_{\text{stim},j} + (160 + i \times 10) \text{ ms}) + \text{bias}$. Red lines (top) and blue lines (bottom) represent *aba* and *bab* responses. They are expected above and below the zero line.

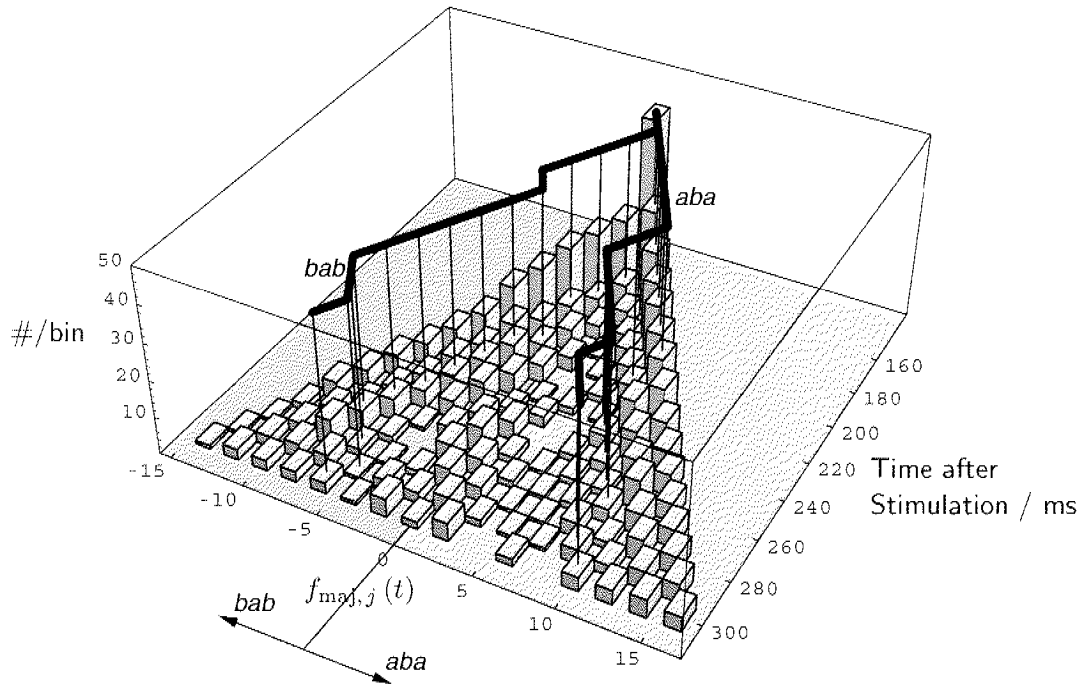


Figure 5.6: Majority-based classification. Plotted is a histogram for each time, $t_i = 160 \text{ ms} + i \times 10 \text{ ms}$, $i = 0 \dots 14$, for which the liquid state has been classified. In the beginning, at $t = 160 \text{ ms}$, all 44 stimulation experiments start at $f_{\text{majority},j}(160 \text{ ms}) = 0$. Two trajectories, one for an *aba* and one for a *bab* stimulation response are plotted separately.

5.4.1.2 Majority-based Classification

Another way to determine the final classification is to neglect the margin of each classification, and to just count the positive versus the negative classifications.

$$f_{\text{maj},j}(t) = \sum_{i=0}^{(160+i \times 10) \text{ ms} \leq t} \text{sign}(c(j, (160 + i \times 10) \text{ ms})) \quad (5.4)$$

with

$$\text{sign}(x) = \begin{cases} 1 & : x \geq 0 \\ -1 & : \text{else} \end{cases} \quad (5.5)$$

$$y_{\text{maj},j} = \text{sign}(f_{\text{maj},j}(300 \text{ ms})) \quad (5.6)$$

Figure 5.6 shows the result for a classification according to a majority vote. Since the trajectories are discrete and would coincide to a large extent in a plot like Figure 5.5, a histogram plot with two exemplary trajectories is given in Figure 5.6.

5.4.2 Anytime Computing with Neuronal Networks

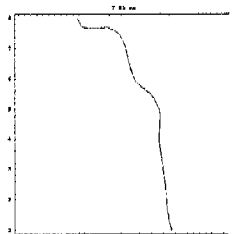
Up to this point, the classification has been binary: either *aba* or *bab*. When the readout is required to provide an output at any time, there is a need for the additional output “neither”. The readout function is substituted by a 2-bit version: $\mathbf{y}(t) : \mathbb{R} \rightarrow \mathbb{B}^2$. Accordingly, two support vector machines, one for each pattern, each consisting of a weight vector, $\mathbf{w}_{i=0,1}$, and a bias, bias_i , are required. For convenience, we define a weight matrix, $\mathbf{W} \in \mathbb{R}^{128 \times 2}$, and a bias vector, $\mathbf{B} \in \mathbb{R}^2$, so that equation 5.1 can be rewritten as

$$c(\mathbf{x}) = \mathbf{W} \cdot \mathbf{x} + \mathbf{B}. \quad (5.7)$$

Two training sets are generated. The respective (*aba* and *bab*) positive training vectors are taken from a time window of 120 to 300 ms after stimulation in steps of 20 ms while the negative examples are taken from 20 ms to 600 ms after stimulation in steps of 40 ms. By choosing the time for the negative examples very long, also response examples for “no stimulation” are included.

For testing, 44 uniformly distributed random time offsets, $\text{off}_{j=1,\dots,44} \in [-10, 10]$ ms, were generated. In the experiment, a inter-stimulus interval of 2.3 s was used. However, because practically all activity ceased at roughly 500 ms after stimulation, we consider only the first 1,200 ms after each stimulation. The liquid states were then sampled during the post-stimulation time span of 10-1,200 ms: $\mathbf{x}_{j,i} = \mathbf{x}(t_{\text{stim},j} + \text{off}_j + i \times 20 \text{ ms})$ with $i = 1 \dots 59$ and $j = 1 \dots 44$, so that a total of 2,596 liquid states was attained for testing.

Figure 5.7 shows the classification of all sampled liquid states. As expected, the majority of liquid states is classified to be neither an *aba* nor a *bab* response.



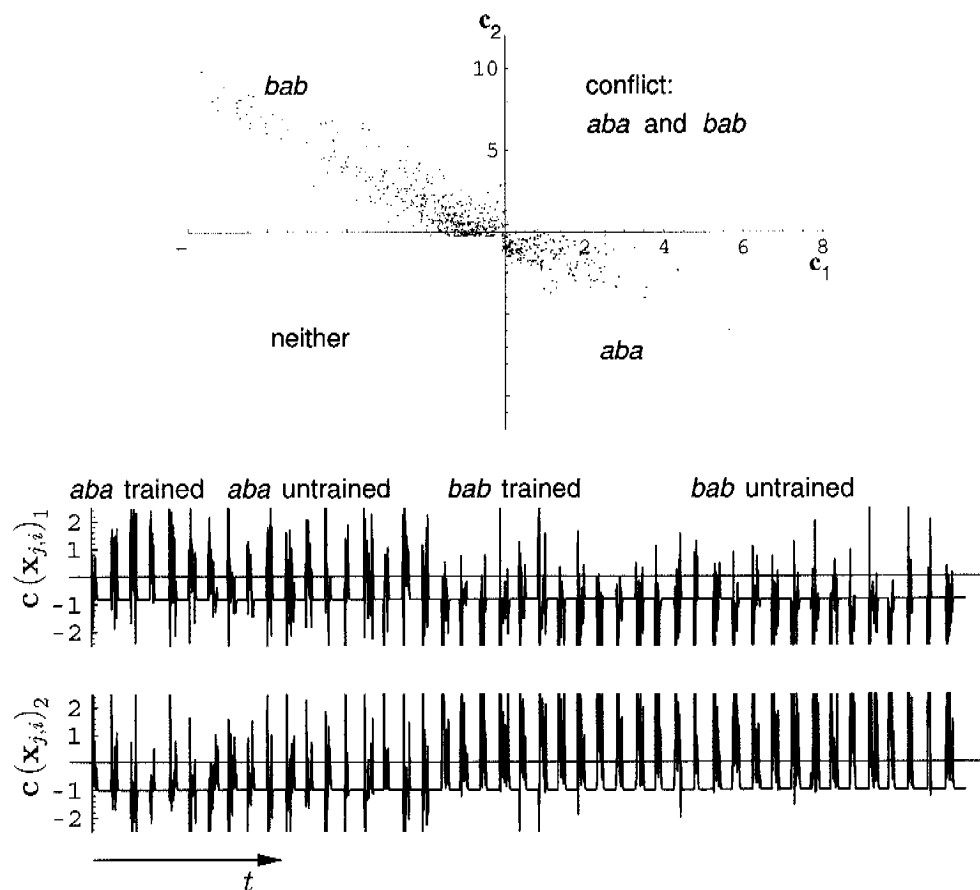


Figure 5.7: Plot of the 2,596 raw classifications results, $\mathbf{c}(\mathbf{x}_{j,i})$. Top: Parametric plot of the points $\mathbf{c}(\mathbf{x}_{j,i}) = \mathbf{W} \cdot \mathbf{x}_{j,i} + \mathbf{B}$. As expected, the majority of liquid states, $\mathbf{x}_{j,i}$, is classified to be neither an *aba* nor a *bab* response. Contradictory classifications, where both classifications are positive, $\mathbf{c}(\mathbf{x}_{j,i})_1 > 0$ and $\mathbf{c}(\mathbf{x}_{j,i})_2 > 0$, can be seen in the upper right quadrant.

Bottom: The same data plotted versus time. The classification is “noisy”.

Since the stimulation responses lasts about 300 ms each, a total of about 30 classifications are obtained per response. In order to convert this high rate of fluctuating classifications (Figure 5.7, bottom) into a more robust output, the results are first low-pass filtered with a time constant of 80 ms and then thresholded. This procedure is a variation of the mean-based classification previously described in section 5.4.1.1. The result can be seen in Figure 5.8: one false negative and three false positives can be distinguished.

5.5 Discussion

Based on the same experiment data, three different liquid state machine configurations have been evaluated for classification of the two different stimulation patterns. All three LSM configurations have in common, that they rely on a leaky integrator to produce the liquid states and that a support vector machine is used to classify the liquid states (section 4.3). The configurations differ in the way the training data and the liquid states are chosen.

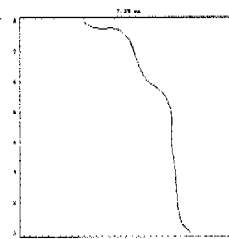
In the first configuration (section 4.3), only one liquid state, exactly at 200 ms post-stimulation, has been used to classify the response per stimulation experiment. 83% of the test set have been correctly classified. In the second configuration (section 5.3), 15 liquid states per stimulation experiment have been used for classification. By averaging the classifications over each stimulation response, 93% have been correctly classified. In the third configuration (section 5.4.2), the liquid state was continuously sampled every 20 ms with an individual random time offset for each stimulation experiment. The liquid state machine automatically detects the occurrence of stimulation patterns and provides a classification output at any time. 93% of the test set has been correctly classified.

5.5.1 Discussion: Anytime versus One-time

When comparing the single-state classification approach that involves only one liquid state per response classification (section 4.3) to the any- and many-time approaches (Chapter 5), two differences are particularly interesting:

- the weight vectors look very different and
- the sensitivities to time jitter in the sampling of the liquid states are very different.

With the single-state classification approach, the alignment of the read out with respect to the stimulation is extremely important. A jitter of 10 ms in the sampling time is sufficient to cause a significant loss of classification quality, whereas the anytime approach shows a mean classification quality that is invariant with



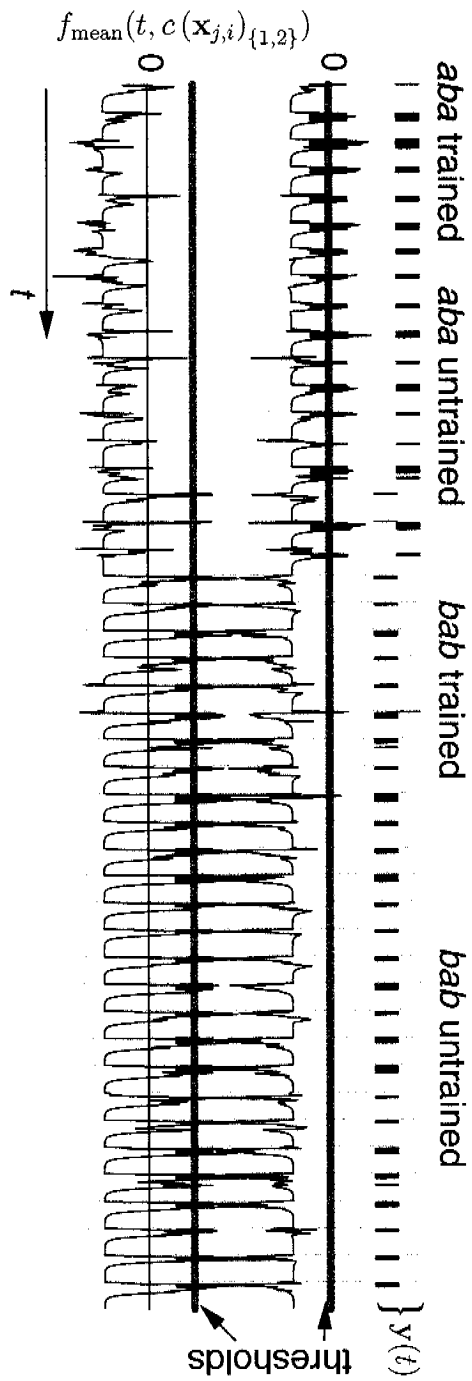


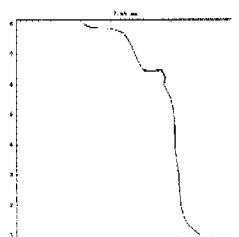
Figure 5.8: Mean-based classification. The data shown in Figure 5.7 is low-pass filtered and compared to thresholds (0 and 1). Stimulations are indicated by long vertical lines. In the upper area of the plot, the final classification output is shown.

respect to the sampling jitter. One kind of interpretation is that there are two different modes according to which the liquid states can be classified:

1. timing-based and
2. synchronicity-based.

When the readout is trained for a precise sampling time, it can rely on processes that happen synchronously to the stimulation. This contrasts to the case of anytime classification, where liquid states from arbitrary times after stimulation are sampled. Therefore, anytime classification can only rely on synchronicity.

This coincides with the observations by Maass et al. [2003], where the classification was found to be robust towards a sinusoidal time wrap of the input pattern.



Chapter 6

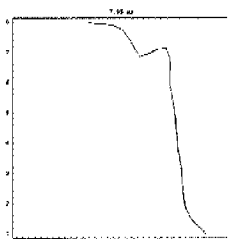
Stimulation

Stimulation is a critical component of neuronal-network-based information processing. There are two issues. First, the stimulation procedure needs to be reliable and may not entail any damage to the neurons. Second, the stimulation artifact needs to be minimized in order to also see the effect of immediately after the stimulation.

6.1 Effective & Gentle Stimulation

For this work, we have exclusively selected the commonly used rectangular, bi-phasic stimulation pulses. Wagenaar et al. [2004] have demonstrated that this way of stimulating is effective. However, there is still no explanation of what exactly is happening. It is not known how the ion channels are activated and how the depolarization reaches the firing threshold. An indicator supporting the assumption that during stimulation something unphysiologic is happening, is the very short stimulation pulse duration. A half pulse length of $100/2 \mu\text{s}$ does probably not suffice to elicit an action potential by opening ion channels and thereby depolarizing the cell to the firing threshold. One possible explanation is that an electroporation of the cell membrane may actually be happening.

Further investigations are needed to render the stimulation more reliable and, especially, to improve the long-term stability of the cells and the stimulation-related effects. One way is to use other stimulation waveforms such as slow voltage ramps as described by Hutzler and Fromherz [2004]. Alternatively other kinds of stimulation can be considered such as stimulation by laser light, either directly, see [Callaway and Yuste, 2002], or indirectly by charge emission, see [Starovoytov et al., 2005].



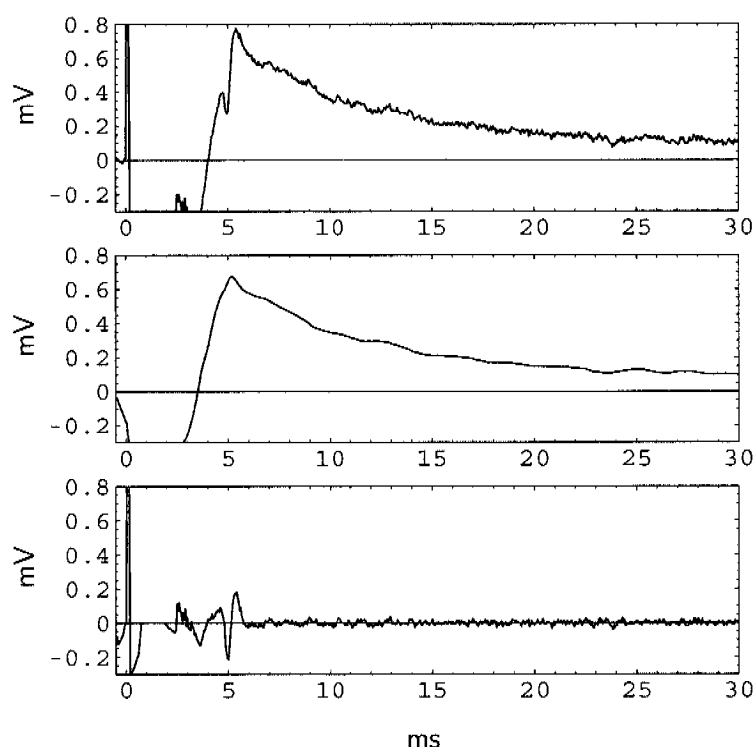


Figure 6.1: Recording of a cortical rat neuron action potential upon stimulation directly from the stimulating electrode using software-based stimulation artifact cancellation. The top graph shows the raw recording. The bi-phasic rectangular stimulation pulse of ± 1 V and 120 μ s duration begins at 0 ms and is followed by a phase of saturation that lasts 3.8 ms. The middle graph shows a low-pass filtered (corner frequency 18 Hz) version of the raw data. The bottom graph shows the high-pass filtered (corner frequency 18 Hz) version of the raw data. The elicited action potential is clearly visible at 5 ms after stimulation.

6.2 Stimulation artifacts

When discussing stimulation artifacts, it is useful to differentiate two kinds of artifacts:

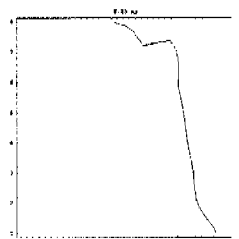
1. the artifact generated on all other, non-stimulated, electrodes by capacitive coupling of the stimulation signal through the culture medium, and
2. the artifact generated on the stimulation electrode itself.

The first kind of artifact is present but does not disturb too much: it is too weak to saturate any filters and so short of time that it can often be neglected.

The second kind, the stimulation artifact on the stimulated electrode, is more severe. After a stimulation pulse on the order of Volts, neuronal activity on the order of a hundred microvolts needs to be recorded. The approach chosen in

this dissertation allows for resetting the readout filters and, thereby, immediately reestablish the operation point of the electronics. The electrodes, however, have their electrochemical equilibrium potential not exactly at the common voltage (2.5 V) but they tend to drift to an offset of roughly 2.5 ± 0.6 V. This offset can be attributed to leakage currents in the CMOS circuitry. In order to quickly reestablish the readout, it is necessary to drive the electrode as close as possible (< 1 mV) back to its equilibrium potential value. Otherwise, the transient signal of the drifting electrode will be too fast to be blocked by the high-pass filter and will, therefore, saturate the downstream amplification and filtering stages. Provided that the stimulation itself does not alter the electrochemical state of the electrode, one possible solution is to sample the DC potential of the electrode before stimulation and copy it back to the electrode after stimulation [Novak and Wheeler, 1988].

In Figure 6.1, an electrode recording of a stimulation is shown. It shows a 3.8 ms long phase of saturation and an action potential recorded 5 ms after stimulation. It also shows that the stimulation artifacts can easily be removed as long as the readout chain is not in saturation. For comparison, for MEA designs without integrated circuitry it can take 40-160 ms after stimulation until it is possible to detect spikes [Wagenaar and Potter, 2004].



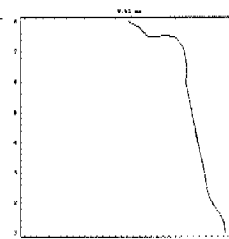
Chapter 7

Conclusion & Outlook

In the framework of this dissertation a system for bidirectional communication with neuronal cultures has been developed. It is based on a CMOS microchip, on top of which the neurons are cultured. Using this system, recordings from a variety of cell types have been presented. The feasibility of information processing with natural neuronal networks (NNN) has been shown on the basis of the acquired experiment data.

Whether or not information processing by means of NNNs will develop into a viable and valuable method of computation, depends largely on biological challenges. Tuning culture parameters like excitability, the ratio of inhibitory to excitatory synapses and thereby modifying the degree of chaotic behavior and the mobility of information, holds great potential. Bicuculline, e.g., a known GABAA receptor antagonist, is an example for a drug that allows to adjust the degree of network activity [Shirakawa et al., 2000]. It can be used to block inhibitory synapses and to thereby change the balance of network inhibition and excitation.

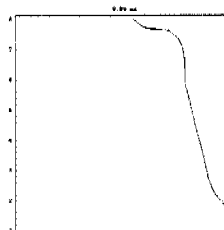
Potential applications for information processing with NNNs will have time-coded input data. Moreover, the time scale of the input to be processed needs to be on the order of the time scale of the NNN activity.



Chapter 8

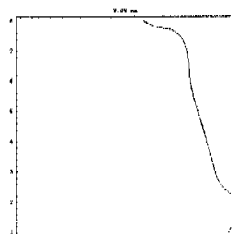
Neurochip Bibliography

- P. J. Ashenden. *The Designer's Guide to VHDL*. Morgan Kaufmann, second edition, 2002.
- P. Auer, H. Burgsteiner, and W. Maass. Reducing communication for distributed learning in neural networks. In *Proc. ICANN 2002*, volume 2415, pages 123–128, 2002.
- B.J. Baker, E.K. Kosmidis, D. Vucinic, C.X. Falk, L.B. Cohen, M. Djuricic, and D. Zeccevic. Imaging brain activity with voltage- and calcium-sensitive dyes. *Cell Mol Neurobiol*, 25(2):245–82, 2005.
- W.H. Baumann, E. Schreiber, G. Krause, A. Podssun, S. Homma, R. Schrott, R. Ehret, I. Freund, and M. Lehmann. Cell monitoring system with multiparametric CMOS sensorchips. *μ TAS*, 2:554 – 556, 2004.
- L. Berdondini, P.D. Van der Wal, O. Guenat, N.F. de Rooij, M. Koudelka-Hep, P. Seitz, R. Kaufmann, P. Metzler, N. Blanc, and S. Rohr. High-density electrode array for imaging in vitro electrophysiological activity. *Biosensors Bioelectronics*, 21:167 – 174, 2005.
- P. Bonifazi and P. Fromherz. Silicon chip for electronic communication between nerve cells by non-invasive interfacing and analog-digital processing. *Adv. Mater.*, 14(17): 1190 – 1193, 2002.
- J.R. Buitengeweg, W.L.C. Rutten, and E. Marani. Extracellular stimulation window explained by a geometry-based model of the Neuron-electrode contact. *IEEE Transactions on Biomedical Engineering*, 49:1591 – 1599, 2002b.
- J.R. Buitengeweg, W.L.C. Rutten, E. Marani, S.K.L. Polman, and J. Ursum. Extracellular detection of active membrane currents in the neuron-electrode interface. *Journal of Neuroscience Methods*, 115:211 – 221, 4 2002c.
- E.M. Callaway and R. Yuste. Stimulating neurons with light. *Curr Opin Neurobiol*, 12:587 – 592, 2002.



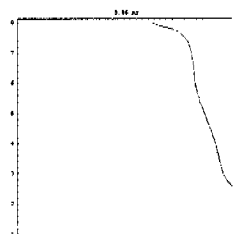
- CESYS GmbH. CESYS USB2 board combining a Xilinx Spartan II, XC2S200 with a Cypress FX-2 chip set. Herzogenaurach, Germany.
- B. V. Chirikov. A universal instability of many-dimensional oscillator systems. *Phys. Rep.*, 52:264–379, 1979.
- Enric Claverol-Tinture and Jerome Pine. Extracellular potentials in low-density dissociated neuronal cultures. *J. Neurosci. Meth.*, 117(1):13–21, May 2002.
- K.S. Colc. Dynamic electrical characteristics of the squid axon membrane. *Arch. Sci. physiol.*, 3:253 – 258, 1949.
- B.D. DeBusschere and G.T.A. Kovacs. Portable cell-based biosensor system using integrated CMOS cell-cartridges. *Biosens. Bioelectron.*, 16:543 – 556, 9 2001.
- A. Dembo and T. Kailath. Model-free distributed learning. *IEEE Transactions on Neural Networks*, 1(1):58 – 70, 1990.
- R.O. Duda and P.E. Hart. *Pattern Classification*. John Wiley & Sons, New York, 2 edition, 2001.
- EpoteK. EpoteK. Billerica, MA, USA. URL <http://www.epotek.com>.
- B. Eversmann, M. Jenkner, F. Hofmann, C. Paulus, R. Brederlow, B. Holzapfl, P. Fromherz, M. Merz, M. Brenner, M. Schreiter, R. Gabl, K. Plehnert, M. Steinhäuser, G. Eckstein, D. Schmitt-Landsiedel, and R. Thewes. A 128 x 128 CMOS biosensor array for extracellular recording of neural activity. *IEEE Journal of Solid-State Circuits*, 38:2306 – 2317, 2003a.
- C. Fernando and S. Sojakka. Pattern recognition in a bucket. In *ECAL 2003*, pages 588 – 597, 2003.
- I.R. Fiete and S.H. Seung. Gradient learning in spiking neural networks by dynamic perturbation of conductances. *Physical Review Letters*, 97(4):048104, 2006. URL <http://link.aps.org/abstract/PRL/v97/e048104>.
- R.V. Florian. Biologically inspired neural networks for the control of embodied agents. 2003. URL <http://www.coneural.org/reports/Coneural-03-03.pdf>.
- W. Franks, I. Schenker, P. Schmutz, and A. Hierlemann. Impedance characterization and modeling of electrodes for biomedical applications. *Biomedical Engineering, IEEE Trans on*, 52(7):1295 – 1302, 2005.
- P. Fromherz. Electrical interfacing of nerve cells and semiconductor chips. *Chemphyschem*, 3:276 – 284, 2002.
- D. Goldenholz. *Liquid Computing: A Real Effect*. Final Project, Boston University, Boston, USA, 2003. URL <http://www.lsm.tugraz.at/papers/Goldenholz-report.pdf>.
- F. Greve. *Micromachined platforms for manipulating and recording from cells*. Ph.D. Thesis, No. 16526, ETH Zurich, Zurich, 2006.

- G. W. Gross, E. Ricske, G. W. Kreutzberg, and A. Meyer. A new fixed-array multi-microelectrode system designed for long-term monitoring of extracellular single unit neuronal activity in vitro. *Neuroscience Letters*, 6:101 – 105, November 1977.
- G.W. Gross, B.K. Rhoades, D.L. Reust, and F.U. Schwalm. Stimulation of monolayer networks in culture through thin-film indium-tin oxide recording electrodes. *Journal of Neuroscience Methods*, 50:131 – 143, 11 1993.
- S. Hafizovic. Ce usb kit, 2005. URL <http://sourceforge.net/projects/ce-usb-kit>.
- S. Hafizovic, F. Heer, F. Greve W. Franks, A. Blau, C. Ziegler, and A. Hierlemann. CMOS bidirectional electrode array for electrogenic cells. In *MEMS 2006*, pages 4 – 7, Istanbul, Turkey, 2006.
- F. Heer. *CMOS-based Microelectrode Array for Communication with Electrogenic Cells*. Ph.D. Thesis, No. 16330, ETH Zurich, Zurich, 2005.
- F. Heer, W. Franks, A. Blau, S. Taschini, C. Ziegler, A. Hierlemann, and H. Baltes. CMOS microelectrode array for the monitoring of electrogenic cells. *Biosensors Bioelectronics*, 20:358 – 366, 9 2004.
- F. Heer, S. Hafizovic, W. Franks, A. Blau, C. Ziegler, and A. Hierlemann. Cmos microelectrode array for bidirectional interaction with neuronal networks. *Solid-State Circuits, IEEE Journal of*, 41:1620 – 1629, 2006a.
- M.O. Heuschkel, M. Fejtl, M. Raggenbass, D. Bertrand, and P. Renaud. A three-dimensional multi-electrode array for multi-site stimulation and recording in acute brain slices. *Journal of Neuroscience Methods*, 114:135 – 148, 3 2002.
- M.L. Hines and N.T. Carnevale. The neuron simulation environment. *Neural Computation*, 9:1179–1209, 1997.
- A.L. Hodgkin and A.F. Huxley. A quantitative description of membrane current and its application to conduction and excitation in nerve. *J. Physiol.*, 117:500 – 544, 1952.
- J.J. Hopfield and C.D. Brody. What is a moment? "Cortical" sensory integration over a brief interval. *Proc Natl Acad Sci U S A*, 97(25):13919 – 24, 2000.
- M. Hutzler and P. Fromherz. Silicon chip with capacitors and transistors for interfacing organotypic brain slice of rat hippocampus. *Eur J Neurosci*, 19(8):2231–8, 2004.
- C.D. James, A.J.H. Spence, N.M. Dowell-Mesfin, R.J. Hussain, K.L. Smith, H.G. Craighead, M.S. Isaacson, W. Shain, and J.N. Turner. Extracellular recordings from patterned neuronal networks using planar microelectrode arrays. *Biomedical Engineering, IEEE Transactions on*, 51:1640 – 1648, 2004.
- M. Jenkner, B. Muller, and P. Fromherz. Interfacing a silicon chip to pairs of snail neurons connected by electrical synapses. *Biological Cybernetics*, 84:239 – 249, 4 2001.



- Y. Jimbo and A. Kawana. Electrical-stimulation and recording from cultured neurons using a planar electrode array. *Bioelectrochemistry and Bioenergetics*, 29:193 – 204, 1992.
- Y. Jimbo and H.P.C. Robinson. Propagation of spontaneous synchronized activity in cortical slice cultures recorded by planar electrode arrays. *Bioelectrochemistry*, 51: 107 – 115, 6 2000.
- Y. Jimbo, N. Kasai, K. Torimitsu, T. Tateno, and H.P.C. Robinson. A system for MEA-based multisite stimulation. *Biomedical Engineering, IEEE Transactions on*, 50:241 – 248, 2003.
- E.R. Kandel, J.H. Schwartz, and T.M. Jessell. *Principles of Neural Science*. McGraw-Hill Companies, 2000.
- M. Kerszberg and J.-P. Changeux. A simple molecular model of neurulation. *BioEssays*, 20:758–770, 1998.
- A.G. Kleber and Y. Rudy. Basic mechanisms of cardiac impulse propagation and associated arrhythmias. *Physiol. Rev.*, 84:431 – 488, 2003.
- G.T.A. Kovacs. Electronic sensors with living cellular components. *Proceedings of the IEEE*, 91:915 – 929, 2003.
- A. Lambacher, M. Jenkner, M. Merz, B. Eversmann, R.A. Kaul, F. Hofmann, R. Thewes, and P. Fromherz. Electrical imaging of neuronal activity by multi-transistor-array (MTA) recording at 7.8 μm resolution. *Applied Physics A-Materials Science & Processing*, 79:1607 – 1611, 2004.
- Michael S Lewicki. A review of methods for spike sorting: the detection and classification of neural action potentials. *Comput. Neural Syst.*, pages 53 – 78, 9 1998.
- W. Maass, T. Natschläger, and H. Markram. Real-time computing without stable states: A new framework for neural computation based on perturbations. *Neural Computation*, 14(30):2531 – 2560, 2002.
- W. Maass, T. Natschlaeger, and H. Markram. A model for real-time computation in generic neural microcircuits. *Advances in Neural Information Processing Systems*, 15:229 – 236, 2003.
- M.P. Maher, J. Pine, J. Wright, and Y.-C. Tai. The neurochip: a new multielectrode device for stimulating and recording from cultured neurons. *Journal of Neuroscience Methods*, 87:45 – 56, 2 1999.
- M. Mahowald and R. Douglas. A silicon neuron. *Nature*, 354:515 – 518, 1991.
- S. Marom and G. Shahaf. Development, learning and memory in large random networks of cortical neurons: lessons beyond anatomy. *Quarterly Reviews Of Biophysics*, 35: 63 – 87, 2 2002.

- S. Martinoia, P. Massobrio, M. Bove, and G. Massobrio. Cultured neurons coupled to microelectrode arrays: circuit models, simulations and experimental data. *IEEE Trans Biomed Eng*, 51(5):859–64, 2004.
- C. C. McIntyre and W. M. Grill. Extracellular stimulation of central neurons: influence of stimulus waveform and frequency on neuronal output. *J Neurophysiol*, 88(4):1592–604, 2002.
- MCS. Multi Channel Systems GmbH. <http://www.multichannelsystems.com>, Reutlingen, Germany.
- P. Mendes. Biochemistry by numbers: simulation of biochemical pathways with Gepasi 3. *Trends Biochem. Sci.*, 22:361–363, 1997.
- Mentor Graphics Corp. Modelsim se plus 6.0a. Wilsonville, Oregon, USA, 2004.
- S.I. Morefield, E.W. Keefer, K.D. Chapman, and G.W. Gross. Drug evaluations using neuronal networks cultured on microelectrode arrays. *Biosens. Bioelectron.*, 15:383–396, 10 2000.
- F.O. Morin, Y. Takamura, and E. Tamiya. Investigating neuronal activity with planar microelectrode arrays: Achievements and new perspectives. *Bioscience and Bioengineering*, 100(2):131–143, 2005.
- E. Neher and B. Sakmann. Single-channel currents recorded from membrane of denervated frog muscle fibres. *Nature*, 260:799–802, 1976.
- Roland Nilson, Johan Björkegren, and Jesper Tegnér. A flexible implementation for support vector machines. *Mathematica Journal*, 10(1), 2006. URL <http://www.mathematica-journal.com/issue/v10i1>.
- J.L. Novak and B.C. Wheeler. Multisite hippocampal slice recording and stimulation using a 32 element microelectrode array. *J Neurosci Methods*, 23(2):149–59, 1988.
- A.L. Obaid, L.M. Loew, J.P. Wuskell, and B.M. Salzberg. Novel naphthylstyryl-pyridium potentiometric dyes offer advantages for neural network analysis. *J Neurosci Methods*, 134(2):179–90, 2004.
- R.H.I. Olsson, D.L. Buhl, A.M. Sirota, G. Buzsaki, and K.D. Wise. Band-tunable and multiplexed integrated circuits for simultaneous recording and stimulation with microelectrode arrays. *Biomedical Engineering, IEEE Transactions on*, 52:1303–1311, 2005.
- J.J. Pancrazio, S.A. Gray, Y.S. Shubin, N. Kulagina, D.S. Cuttino, and K.M. Shaffer. A portable microelectrode array recording system incorporating cultured neuronal networks for neurotoxin detection. *Biosensors Bioelectronics*, 18(11):1339–1347, 2003.
- J. Pelt, P.S. Wolters, M.A. Corner, W.L.C. Rutten, and G.J.A. Ramakers. Long-term characterization of firing dynamics of spontaneous bursts in cultured neural networks. *IEEE Transactions on Biomedical Engineering*, 51(11):2051–2062, 2004.



- Z.A. Peterlin, J. Kozloski, B.Q. Mao, A. Tsiola, and R. Yuste. Optical probing of neuronal circuits with calcium indicators. *Proc Natl Acad Sci U S A*, 97(7):3619–24, 2000.
- S. M. Potter. Distributed processing in cultured neuronal networks. *Prog Brain Res*, 130:49–62, 2001.
- K. Ramasubramanian and M. S. Sriram. A comparative study of computation of lyapunov spectra with different algorithms. *Phys. D*, 139(1-2):72–86, 2000. ISSN 0167-2789. doi: [http://dx.doi.org/10.1016/S0167-2789\(99\)00234-1](http://dx.doi.org/10.1016/S0167-2789(99)00234-1).
- C. Rasche and R. Douglas. An improved silicon neuron. *Analog Integrated Circuits and Signal Processing*, 23(3):227 – 236, 2000.
- W.L.C. Rutten. Selective electrical interfaces with the nervous system. *Annual Review Of Biomedical Engineering*, 4:407 – 452, 2002.
- Tetsuo Shirakawa, Sato Honma, Yumiko Katsuno, Haruhisa Oguchi, and Ken ichi Honma. Synchronization of circadian firing rhythms in cultured rat suprachiasmatic neurons. *European Journal of Neuroscience*, 12(8):2833–2838, 2000. doi: 10.1046/j.1460-9568.2000.00170.x.
- A. Starovoytov, J. Choi, and H.S. Seung. Light-directed electrical stimulation of neurons cultured on silicon wafers. *J Neurophysiol*, 93(2):1090–8, 2005.
- A. Stett, U. Egert, E. Guenther, F. Hofmann, T. Meyer, W. Nisch, and H. Haemmerle. Biological application of microelectrode arrays in drug discovery and basic research. *Analytical and Bioanalytical Chemistry*, V377:486–495, October 2003b. doi: 10.1007/s00216-003-2149-x.
- Inc. Synplicity. Synplify 8.1. California, USA, 2005.
- Makoto Taketani and Michel Baudry. *Advances in Network Electrophysiology*. Springer, New York, 2006.
- S. Taschini, M. Emmenegger, H. Baltes, and J. G. Korvink. Smart enumeration in C++: Virtual construction, message dispatching and tables. *Software-Practice and Experience*, 29:67 – 76, 1999.
- Trolltech, Inc. Qt 4.2. Oslo, Norway, 2007.
- USB Implementers Forum. USB specification, 2005. URL <http://www.usb.org/developers/docs>.
- M. Voelker and P. Fromherz. Signal transmission from individual mammalian nerve cell to field-effect transistor. *Small Journal.*, 1:1–4, 2005.
- D. Wagenaar, J. Pine, and S. Potter. Searching for plasticity in dissociated cortical cultures on multi-electrode arrays. *Journal of Negative Results in BioMedicine*, 5(1):16, 2006b. doi: 10.1186/1477-5751-5-16.

- D.A. Wagenaar and S.M. Potter. A versatile all-channel stimulator for electrode arrays, with real-time control. *Journal of Neural Engineering*, 1(1):39 – 45, 2004. URL <http://stacks.iop.org/1741-2552/1/39>.
- D.A. Wagenaar, J. Pine, and S.M. Potter. Effective parameters for stimulation of dissociated cultures using multi-electrode arrays. *J Neurosci Methods*, 138(1-2): 27–37, September 2004. doi: 10.1016/j.jneumeth.2004.03.005.
- D.A. Wagenaar, R. Madhavan, J. Pine, and S.M. Potter. Controlling bursting in cortical cultures with closed-loop multi-electrode stimulation. *J. Neurosci.*, 25:680 – 688, 2005.
- C. Windeck. Gestählter Goliath. *c't Magazin für Computer Technik*, 16:106 – 111, 2006.
- K.D. Wise, D.J. Anderson, J.F. Hetke, D.R. Kipke, and K. Najafi. Wireless implantable microsystems: high-density electronic interfaces to the nervous system. *Proceedings of the IEEE*, 92:76 – 97, 2004.
- Inc. Wolfram Research. Mathematica 5.2. Champaign, USA, 2005.
- XFAB. 0.6 μm process, triple-metal, double-polysilicon CMOS process. Dresden, Germany.
- Xilinx. Xilinx ISE 7.1. San Jose, California, USA, 2005.

Part II

Single-Chip Atomic Force Microscope

Seite Leer /
Blank leaf

Chapter 9

Introduction

This part of the thesis reports on work in the context of the single-chip AFM project realized from 2002 to 2003. Based on a cantilever design available from previous work [Lange et al., 2002a], a new chip including digital, analog and mechanical components was devised a core team of three people:

- Diego Barrettino: analog circuits
- Sadik Hafizovic: digital on-chip circuits & controller, integration, measurements
- Tormod Volden: cantilevers, in-house micromachining

9.1 Structure

In the following the reprints of the two resulting, most relevant papers are given:

1. *CMOS Monolithic Mechatronic Microsystem for Surface Imaging and Force Response Studies*
D. Barrettino, S. Hafizovic, T. Volden, J. Sedivy, K.-U. Kirstein, and A. Hierlemann, *Journal of Solid-State Circuits* 2005, 40(4), 951-959 and
2. *Single-Chip Mechatronic Microsystem for Surface Imaging and Force Response Studies*
S. Hafizovic, D. Barrettino, T. Volden, J. Sedivy, K.-U. Kirstein, O. Brand, and A. Hierlemann *Proc Natl Acad Sci USA*. 2004 Dec 7, 101(49), 17011-17015.

The first paper describes the circuitry of the integrated atomic-force microscope in great detail. The second paper focuses on applications and measurement results and is intended for a broader audience. Concise information on the integrated cantilevers is given in a separate book chapter [Hafizovic et al., 2007].

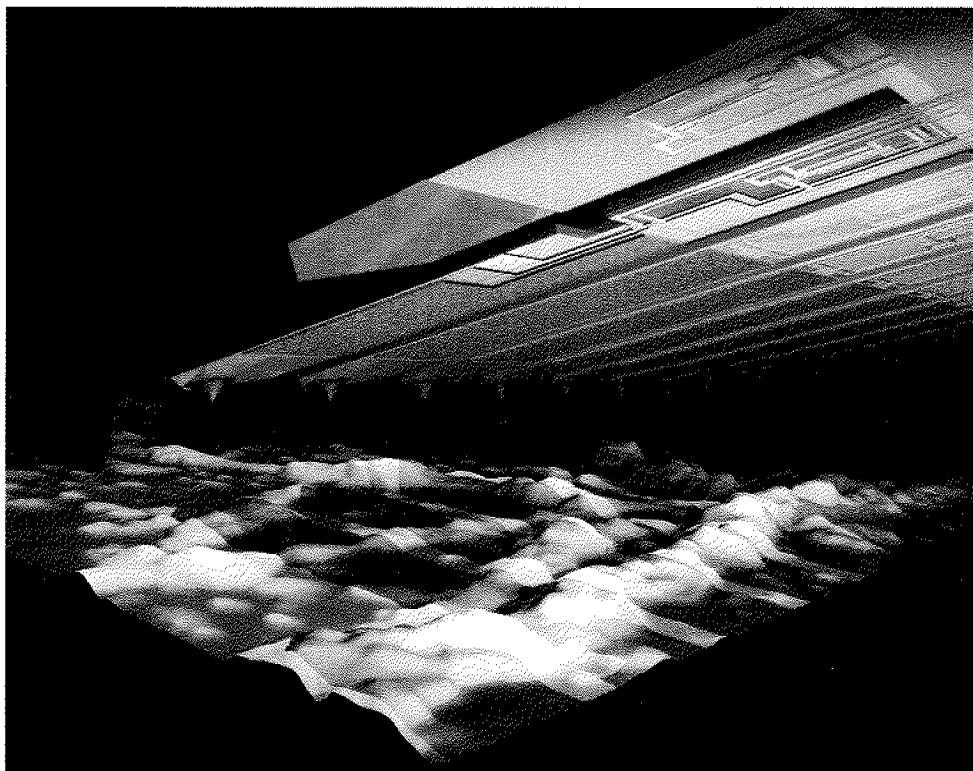


Figure 9.1: Artistic rendering of the 10-cantilever AFM array. The first, left, short cantilever is a reference cantilever. Then, there are 10 scanning cantilevers, 4 of which can be simultaneously operated. Also visible are the diffused piezo resistors (yellow) in a Wheatstone configuration near the base of each cantilever for force detection and the thermal actuators (aluminum; brown) on top of the diffused heating resistors (yellow) that occupy half of the area of the 500 μm -long cantilevers. The sample represents original data, see Figure 11.6 on page 103.

Chapter 10

CMOS Monolithic Mechatronic Microsystem for Surface Imaging and Force Response Studies

D. Barrettino, S. Hafizovic, T. Volden, J. Sedivy,
K.-U. Kirstein, and A. Hierlemann
Journal of Solid-State Circuits 2005, 40(4), pp. 951-959

Physical Electronics Laboratory, ETH Zurich, Switzerland

Abstract — We report on a standalone single-chip (7 mm×10 mm) atomic force microscopy (AFM) unit including a fully integrated array of cantilevers, each of which has individual actuation, detection, and control units so that standard AFM operations can be performed only by means of the chip without any external controller. The system offers drastically reduced overall size and costs and can be fabricated in standard CMOS technology with some post-CMOS micromachining steps to form the cantilevers. Full integration of microelectronic and micromechanical components on the same chip allows for controlling and monitoring all system functions. The on-chip circuitry notably improves the overall system performance. Circuitry includes analog signal amplification and filtering stages with offset compensation, analog-to-digital converters, digital proportional-integral-derivative deflection controllers, sensor-actuator compensation filters, and an on-chip digital interface for data transmission. The microsystem characterization evidenced a vertical resolution of better than 1 nm and a force resolution of better than 1 nN as shown in the measurement results. This CMOS monolithic AFM microsystem allows for precise and fully controlled mechanical manipulation in the nanoworld.

10.1 Introduction

New measurement, metrology, and imaging techniques have been pivotal to the rapid development of many branches of science such as materials science, microelectronics and microbiology to name but a few. The invention of the scanning-tunneling microscope (STM) by Binnig and Rohrer in 1982 [Binnig and Rohrer, 1982] and, in particular, the invention of the atomic force microscope (AFM) in 1986 [Binnig et al., 1986] have established a basis for many findings in various scientific areas over the last years. The AFM has evolved at an exceptional speed from a laboratory prototype to a commercial instrument, see, e.g., Inc., and many other scanning probe techniques have been developed, which will not be further mentioned here [Meyer and Heinzelmann, 1995]. The AFM has been used to measure forces during stretching of DNA strands [Rief et al., 1999] and rupture forces of single covalent bonds [Grandbois et al., 1999], and AFM can be operated in liquids, which enables its use in biological applications [Lehenkari et al., 2000, You et al., 2000, Bowen et al., 1998]. The AFM also can be used to perform surface manipulations such as lithographic fabrication of a transistor [Minne et al., 1996, 1999], or AFM-based storage [Mamin et al., 1999, Vettiger et al., 2002].

Commercially available AFM instruments are rather bulky and have a low throughput due to the serial nature of the involved scanning process and the limited scanning range, which renders the investigation of larger samples rather laborious. The detection of the cantilever deflection is done mostly by means of a laser, which is costly and makes the adjustment and cantilever exchange very time-consuming, in particular, when operating in vacuum environment. To overcome these limitations, AFM probes have been developed with integrated detection schemes such as capacitive [Brugger et al., 1992], piezoelectric, or piezoresistive schemes [Minne et al., 1999, Linnemann et al., 1995, Gotszalk et al., 1998, Jumpertz et al., 1998, Thaysen et al., 2000], and high-speed scanning systems that rely on arrays of cantilevers featuring piezoelectric excitation and piezoresistive/piezoelectric readout [Minne et al., 1999, 1998b,a, Kim et al., 1998]. All of those systems, however, require a larger set of desktop equipment since no integrated electronics or functions are provided.

Earlier work on integrated AFM instruments included a chip featuring up to ten cantilevers and analog time-multiplexed actuation. Readout of the cantilevers featured rather limited force and vertical resolution and an external deflection controller was used [Lange et al., 2002a].

In this paper, we will give a detailed description of a standalone CMOS monolithic AFM unit including a fully-integrated array of cantilevers, each of which has its individual actuation, detection, control, and amplification as well as its individual offset compensation, as presented in [Barrettino et al., 2003, Barrettino, 2004, Hafizovic et al., 2004]. We will show constant-force measurements and multiple force-distance measurements performed without any external controllers. The

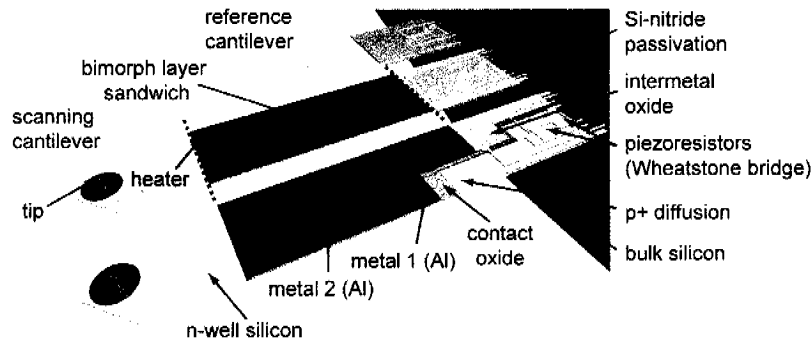


Figure 10.1: Three-dimensional schematic of three cantilevers (two scanning cantilevers and one shorter reference cantilever) showing the different CMOS layers as they are used for the various cantilever components (actuation: bimorph and heater; detection: piezoresistive Wheatstone bridge). A color version of this figure is given on page 98.

cantilevers can be moved and precisely controlled within a range of 0.5 to 6 μm (at 0.5 to 6 nm resolution) so that the chip has only to be brought within a distance of maximal 6 μm to a surface. Only an x-y-scanning stage is necessary for imaging or lateral scanning by using one or several of the individually controlled cantilevers. The individual cantilever control is of paramount importance in using multicantilever arrays since there is always a slight tilt between the chip plane determining the cantilever positions and the sample surface plane so that some cantilevers are closer to the surface than others.

Parallel scanning of multiple cantilevers, therefore, is facilitated by implementing a fast and simultaneous individual force feedback for each scanning cantilever in a closed loop. On the other hand, time-domain multiplexing of digital circuits would reduce the area of the chip and will be implemented in future versions of this microsystem.

This paper is organized as follows. Section 10.2 describes the microcantilever array and the post-CMOS fabrication steps. Section 10.3 contains details of the circuit design. In Section 10.4, the experimental results are presented, and Section 10.5 concludes the paper.

10.2 Microcantilever Array

A detailed 3-dimensional schematic of the scanning microcantilever is shown in Figure 10.1, and the micrograph of the cantilever array is shown in Figure 10.2. The cantilever array consists of twelve cantilevers. There are ten scanning cantilevers, which are 500 μm long, 85 μm wide, and 5 μm thick, and two reference cantilevers, which are 250 μm long, 85 μm wide, and 5 μm thick. The spacing among the cantilevers is 25 μm , and they feature a force constant of 1 N/m. The

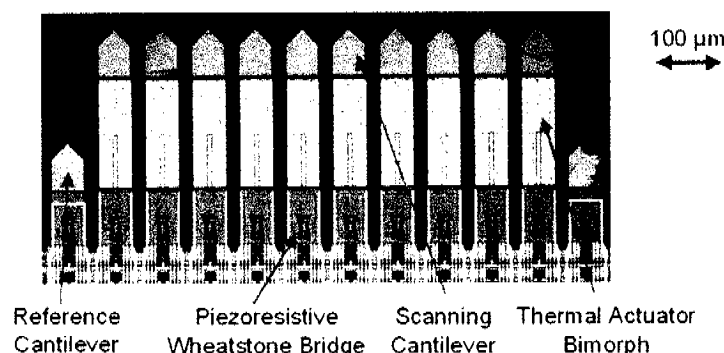


Figure 10.2: Close-up of the cantilever area. The chip features twelve cantilevers, ten of which can be potentially used for scanning (500 μm long, 85 μm wide). Only the four cantilevers in the center are connected to the circuitry. The two shorter ones at the flanks (250 μm long) serve as references.

mechanical resonance frequency of the cantilevers is around 43 kHz, and the Q factor is around 400.

The cantilevers of this microsystem are deflected by making use of the bimorph effect, i.e., the different thermal expansion coefficients of silicon and the aluminum layers (metal 1 and metal 2 of the CMOS process, see Lange et al. [2002a]). Upon heating the layer sandwich, the cantilever bends, and the bending can be precisely controlled through the heating current passing through the p+ diffused heating resistor. The nominal resistance of the p+ diffusion heater is 1.3 k Ω .

To detect the cantilever deflection caused by force exerted on the tip, four p+ diffused piezoresistors (resistors that change their resistance upon mechanical stress), which have a nominal resistance of 1 k Ω , are arranged at the cantilever base in a Wheatstone-bridge configuration [Gotszalk et al., 1998, Jumpertz et al., 1998, Lange et al., 2002a]. Two of them are parallel oriented to the cantilever axis, and two of them are oriented perpendicularly in order to achieve a maximum signal upon deflection [Lange et al., 2002a]. The reference cantilevers feature only the Wheatstone bridge for deflection detection, and they are shorter than the scanning cantilevers to prevent contact with the sample surface. The reference cantilevers are used for offset compensation of the Wheatstone bridge.

The integrated circuitry (described in more detail in Section 10.3) performs the microcantilever deflection control in the vertical direction, as well as offset compensation and amplification of the signals from the piezoresistive Wheatstone bridge.

The cantilevers are fabricated in a standard 0.8 μm CMOS process [AMS]. The cantilever structures are then formed in post-CMOS micromachining steps. First, silicon membranes are created by anisotropic silicon wet-etching using a potassium-hydroxide (KOH) solution applied to the backside of the wafer. An electrochemical etch stop technique provides a defined and uniform thickness of

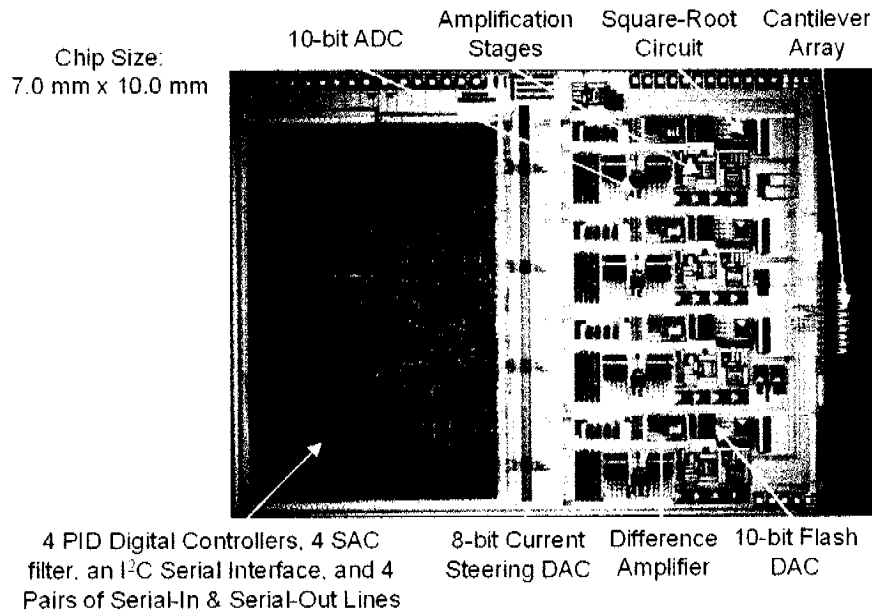


Figure 10.3: Micrograph of the CMOS monolithic AFM microsystem.

the membrane, which consists of the n-well layer of the CMOS process [Müller et al., 2000].

Front-side reactive ion etching (RIE) or wet etching is afterwards used to locally remove oxide layers from the cantilever tip, then front-side RIE is used again to release the cantilevers.

Finally, a silicon nitride tip (radius 10-20 nm) is mounted at the cantilever end [Lange et al., 2002a].

10.3 Circuit Design

Figure 10.3¹ shows a micrograph of the CMOS monolithic AFM microsystem, which is fabricated in a standard 0.8 μm CMOS process [AMS] and features a die size of 7 mm x 10 mm. A simplified block diagram of the main units is shown in Figure 10.4.

The microsystem has a differential mixed-signal architecture. The cantilever deflection signal coming from the Wheatstone bridge is amplified, filtered and translated into the digital domain. When the biasing of the Wheatstone bridge is 5 V, the deflection sensitivity is 5 $\mu\text{V}/\text{nm}$, which in combination with the spring constant of 1 N/m, corresponds to a force sensitivity of 5 $\mu\text{V}/\text{nN}$. The offset voltage of the Wheatstone bridge, mainly attributed to residual stress from layer deposition, is around 50 mV. The maximum offset voltage that the circuitry can handle is 100 mV.

¹A better chip photo can be found in Figure 11.3 on page 99.

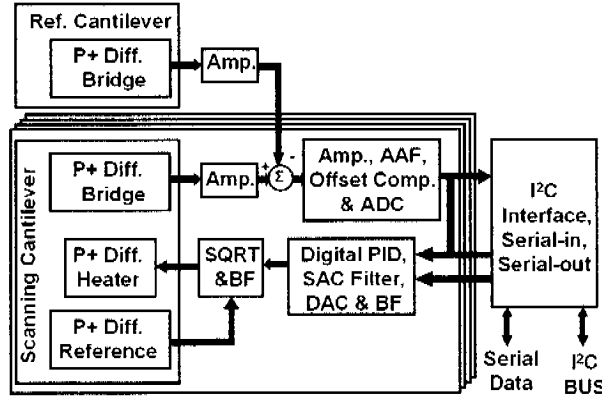


Figure 10.4: Simplified block diagram of the chip architecture. The four frames indicate that these subunits are repeated for each active cantilever, i.e., four times.

All control operations are handled by the on-chip digital controllers, which issue actuation signals to the heaters on the cantilevers via the digital-to-analog converters.

The microsystem electronics mainly consists of four analog blocks, which are connected to the four central scanning cantilevers, and a common digital block. Due to chip area constraints, only four readout and digital processing channels have been realized on-chip to provide a proof of concept. The integration of ten readout-channels can be realized by a redesign in CMOS technology with smaller feature size.

Each analog block includes two fully differential programmable amplifiers and a fixed amplifier for the readout of the small signals coming from the Wheatstone bridge, a difference amplifier for the cancellation of the Wheatstone bridge offset voltage, a pseudo-differential 10-bit successive-approximation analog-to-digital converter (ADC), a programmable 8-bit current steering digital-to-analog converter (DAC) for fine tuning of the remaining offset of the Wheatstone bridge, a geometric mean circuit (or square-root circuit) for the linearization of the relationship between the voltage-drop across the heater and its power dissipation, and a flash DAC for the actuation signal to the heater.

The digital block includes eight programmable infinite-impulse-response (IIR) filters, each with 6 coefficients. Four are configured to act as proportional-integral-derivative (PID) controllers, and four are configured to act as sensor-actuator compensation (SAC) filters. The implementation of each IIR filter is given by equation 10.1,

$$A_i = A_{i-1} + \alpha_1 A_{i-1} + \alpha_2 A_{i-2} + \beta_1 D_{i-1} + \beta_2 D_{i-2} + \gamma R \quad (10.1)$$

where γ , β and α are 11-bit signed fixed-point coefficients in the range (-1,1), D_i the deflection of the cantilever (filter input), A_i the actuation signal (filter output), and R the reference signal (filter input). The filter stores two previous deflection and actuation values. The calculation is performed on a 16-bit

Table 10.1: Register bank of the CMOS monolithic AFM microsystem.

Word	Bits	Description
0-4	15-0	multipurpose bits: analog switches for offset compensation and calibration, loop selection (on-chip or off-chip controller), etc.
5	15-0	delay between streamed samples (in clock cycles) for the I ² C and dedicated serial lines.
6	15-0	sampling period of the digital filters.
7-10	10-0	reference force or deflection of 4 cantilevers.
11-16	9-0	6 PID coefficients: $(\alpha_i, \beta_i, \gamma)$ - i.e., one set of coefficients for all cantilevers.
17-40	10-0	24 SAC coefficients: $(\alpha_i, \beta_i, \gamma)$ - i.e., one set of coefficients per cantilever.

multiplier-accumulator (MAC), and the result is clipped to 10 bits. One multiplication and addition are done per clock cycle. The clock frequency is 1 MHz.

Sensor-actuator compensation (SAC) filters compensate for the sensor-actuator crosstalk, which refers to the pseudo force signal attributed to heat generated by the heater on the same cantilever. This cross-talk is proportional to the actuation signal on the order of some percents (between 5% and 20%). In open-loop mode (force distance measurements), where actuation does not depend on the recorded force values, this is acceptable especially since it is easy to remove in data post-processing. In closed-loop mode (surface imaging), however, sensor-actuator cross-talk needs to be eliminated to prevent the crosstalk component from exceeding the preset contact-force. This holds in particular when the contact forces are programmed to values as small as 5 nN. The crosstalk is eliminated by subtracting a scaled and low-pass-filtered version of the actuation signal from the force signal.

The digital block also includes four pairs of dedicated serial lines for testing and cantilever deflection readout, and an inter-IC (I²C) serial interface to handle the programming of the filter coefficients and programmable switches. Table 10.1 summarizes the contents of the register bank in the digital block.

The computing power of the eight IIR filters in the digital block amounts to 16 million arithmetic operations per second, which is one of the highest values ever realized in CMOS-based micromechanical systems. These dampening IIR filters implement oversampling and thereby noise suppression, which allows for precise cantilever position control. For example, for a total deflection range of 1 μm , a resolution of 1 nm has been achieved, which can be further improved to 0.5 nm by lowering the overall deflection range to 0.5 μm .

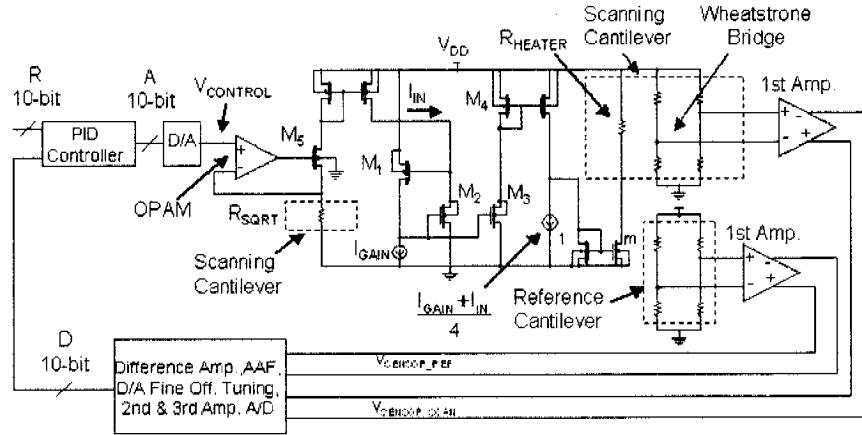


Figure 10.5: Deflection-control loop and square-root circuit of the CMOS monolithic AFM microsystem.

10.3.1 Deflection Control Loop

The microsystem has four deflection control loops, which are connected to the four central scanning cantilevers. One of the control loops is shown in Figure 10.5. After the first amplifier stage, the offset voltage of the piezoresistive Wheatstone bridge in the scanning cantilever ($V_{\text{SENSOR_SCAN}}$) is compensated by subtracting the offset voltage of the reference cantilever structure ($V_{\text{SENSOR_REF}}$). The remaining offset can be compensated by another compensation stage, which enables an individual adjustment of the DC value of each readout channel with 8-bit resolution by means of a programmable digital-to-analog converter (DAC). The overall gain of the analog signal processing is programmable from 18 dB to up to 44 dB to cover the whole range of expectable forces in various applications.

The amplified force signal is converted to the digital domain by a 10-bit pseudodifferential successive-approximation analog-to-digital converter (ADC). The pseudodifferential ADC is a slight modification of the successive-approximation ADC from the analog library of Austriamicrosystems [Barrettino, 2004, AMS], the comparator of which was replaced by a differential comparator, and to which a second DAC was added.

The inputs to the digital PID deflection controller are the feedback signal from the cantilever deflection (D) and the reference signal (R). The output of the digital deflection controller (A) is the input to the SAC filter and the input to the 10-bit flash DAC. The output of the 10-bit flash DAC (V_{CONTROL}) is linearized by the square-root circuit. The linearization of the quadratic relationship between V_{CONTROL} and the power dissipated by the p+ diffusion heater (R_{HEATER}) facilitates the calculation of optimal parameters for the digital PID deflection controller. The linearization, described in ref. [Barrettino et al., 2003], is implemented with a voltage-to-current converter (OPAM, M5, and R_{SQRT}) and a geometric mean circuit or square-root circuit (transistors M1...M4). The control

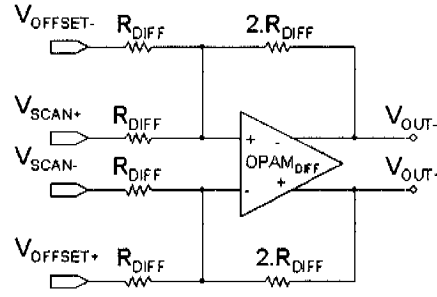


Figure 10.6: Schematic of the difference amplifier.

voltage drives the voltage-to-current converter. The p+ diffusion reference resistor (R_{SQRT}) used in the converter is located on the cantilever in order to be at the same temperature as R_{HEATER} . This configuration eliminates the temperature dependence of the proportionality coefficient between V_{CONTROL} and the power dissipated by R_{HEATER} (P_{HEATER}), given by equation (10.2):

$$P_{\text{HEATER}} = I_{\text{GAIN}} \frac{m^2}{4} \frac{R_{\text{HEATER}}}{R_{\text{SQRT}}} V_{\text{CONTROL}} \quad (10.2)$$

where I_{GAIN} is the bias current of the square-root circuit, m is the transistor ratio of the output current source, and R_{SQRT} is a p+ diffusion resistor located in the scanning cantilever with a nominal resistance of 5.5 k Ω .

The bandwidth of the deflection control loop is around 3 kHz, mainly limited by the cantilever thermal time constant (around 1 ms) and the maximum heating current (electrothermal actuation).

A simplified schematic of the difference amplifier is shown in Figure 10.6. The fully differential operational amplifier (OPAM_DIFF) is a folded-cascode amplifier with buffered outputs [Laker and Sansen, 1994]. The polysilicon resistor (R_{DIFF}) has a nominal resistance of 20 k Ω .

The binary-weight output current of the programmable 8-bit current steering DAC [Baker et al., 1998] flows through a polysilicon resistor of the RC anti-aliasing filter (AAF). Then, the programmable voltage drop across the polysilicon resistor compensates for the remaining offset voltage of the Wheatstone bridge. The voltage drop is temperature-independent since the output current of the DAC is inversely proportional to the resistance of a polysilicon resistor.

10.3.2 Fully Differential Amplifiers

The first amplification stage consists of a fully differential programmable amplifier with source degeneration [Rijns, 1996]. A simplified schematic of the amplifier is shown in Figure 10.7. The maximum (A_{MAX}) and minimum (A_{MIN}) gains of the amplifier are given by equations (3) and (4):

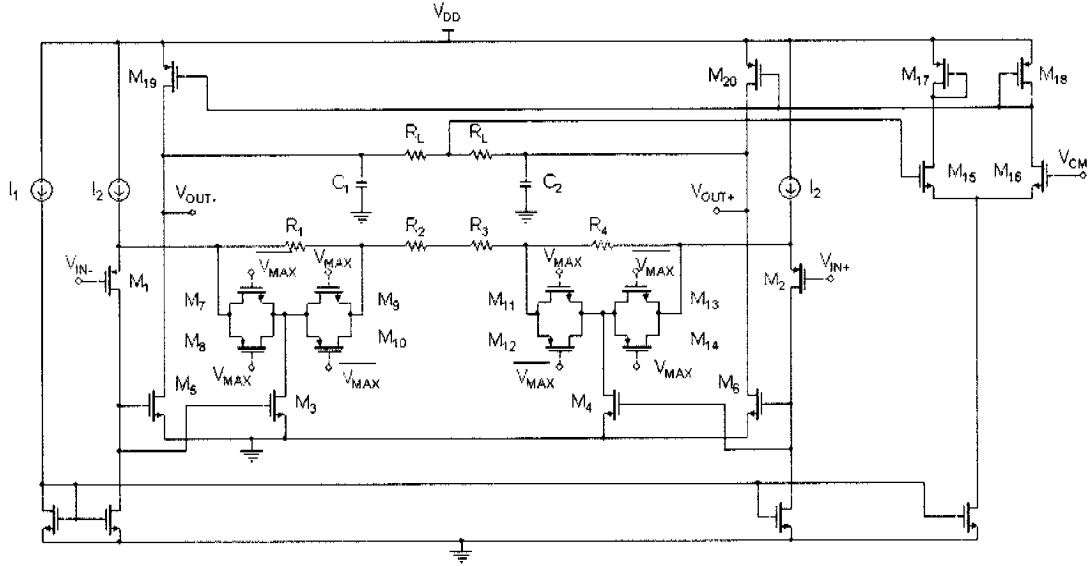


Figure 10.7: Schematic of the fully differential programmable amplifier.

Table 10.2: Experimental results of the first amplification stage.

Parameter	Value
Power Supply	5 V
Current consumption	400 μ A
Gain	18 or 24 dB
Input-referred noise (0.01 Hz to 10 kHz)	< 4 μ V
Unity-gain bandwidth	4.5 MHz
Input-referred offset	< 5 mV
Linear input range	0.5 V < V_{in} < 3.8 V

$$A_{MAX} \simeq \frac{2 \times R_L}{R_2 + R_3} \quad (10.3)$$

$$A_{MIN} \simeq \frac{2 \times R_L}{R_1 + R_2 + R_3 + R_4} \quad (10.4)$$

The programmable degeneration resistor is implemented with four CMOS switches (transistors $M_7 \dots M_{14}$) and four resistors (resistors $R_1 \dots R_4$). The load resistors (resistors R_L) have a nominal resistance of 26 k Ω , whereas the degeneration resistors (resistors $R_1 \dots R_4$) have a nominal resistance of 1.6 k Ω . When the control signal (V_{MAX}) is 0 V, the first amplification stage has a gain of 8, otherwise when V_{MAX} is 5 V the first amplification stage has a gain of 16. The common mode feedback (transistors $M_{15} \dots M_{20}$) adjusts the common mode output to 2.5 V. The experimental results of the first amplification stage are summarized in Table 10.2.

The second amplification stage also consists of a fully differential programmable amplifier with source degeneration. The load resistors have a nominal resistance

of 26 k Ω , whereas the degeneration resistors R_1 and R_4 have a nominal resistance of 23.4 k Ω and resistors R_2 and R_3 have a nominal resistance of 2.6 k Ω . Then, using equations (3) and (4), the gains are 10 and 1.

The third amplification stage is based on a fully differential balanced-output operational amplifier. The details of an implementation of this operational amplifier can be found in ref. [Banu et al., 1988]. The operational amplifier is connected in noninverting configuration with two resistor pairs of 2 k Ω and 20 k Ω , resulting in a fixed gain of 10.

In summary, the maximum amplification gains are used for high-resolution deflection control within a small deflection range (around 0.5 μm). The minimum amplification gains are used for low-resolution deflection control within a large deflection range (around than 6 μm). The gains should be selected according to the application.

10.4 Experimental Results

Two prototype applications were selected to show the performance of the CMOS monolithic AFM microsystem: (a) surface imaging and (b) force distance measurements. For operating the CMOS monolithic AFM microsystem, only a simple printed circuit board (PCB) is necessary to provide some reference voltages. There is no need for external signal processing capacity either for closed-loop imaging or open-loop force 10 measurements. All operations can be performed on-chip. The digital interface connects the microsystem to a computer for data capturing and visualization.

10.4.1 Surface Imaging

For surface imaging, the PCB with the CMOS monolithic AFM microsystem is mounted on the x-y stage of a Nanoscope III, Digital Instruments, USA. The microsystem was operated in contact mode. The on-chip deflection controller measures the force acting on the cantilever and keeps it constant while the tip is scanned over the surface. Height information of the scanned sample is obtained from the actuation signal that is required to keep the cantilever force constant. Figure 10.8 shows a high-resolution small-range line scan recorded at 3 $\mu\text{m/s}$ over one 18 nm step. The maximum possible scanning speed of the monolithic system is approximately 1 mm/s (mainly limited by the cantilever thermal time constant). The maximum readout rate of the force signal is 100 kHz, which allows for averaging multiple values per data point for better noise suppression. Each displayed data point in Figure 10.8 represents the time average of 300 values. The measured error signal indicates an excellent tracking of the surface topography. A vertical resolution of better than 1 nm has been achieved (Figure 10.8).

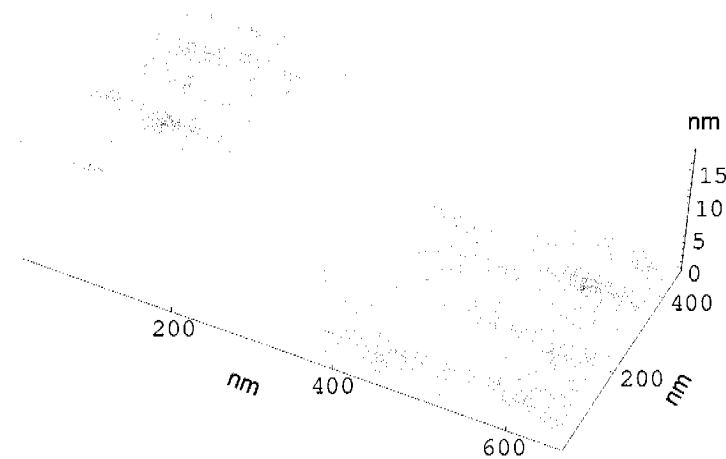


Figure 10.8: High-resolution line-scan over one 18 nm step. Each displayed data point represents the time average of 300 values. The vertical resolution is better than 1 nm.

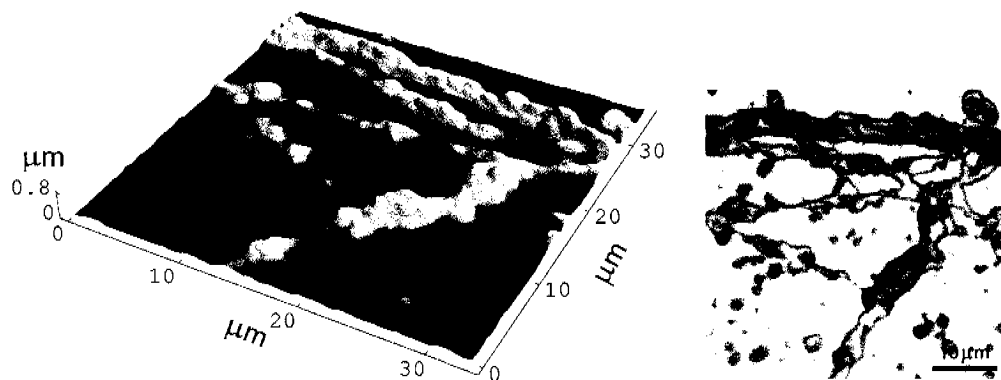


Figure 10.9: Left: scan of a biological sample at a force of 10 nN and a scanning speed of 100 $\mu\text{m/s}$ representing a network of dried-out chicken neurites on a silicon oxide surface after fixation². Right: light-microscope image of the same area of a network of dried-out chicken neurites for comparison [Hafizovic et al., 2004].

A larger area-scan of a biological sample at a force of 10 nN and a scanning speed of 100 $\mu\text{m/s}$ is shown in Figure 10.9, left representing a network of dried-out chicken neurites on a silicon oxide surface after fixation². A light-microscope image is given for comparison (Figure 10.9, right). The force exerted on a sample can be kept as low as 5 nN so that also soft samples can be imaged. Tapping-mode operation can be also realized with the CMOS monolithic AFM microsystem, but this has not yet been tested.

²A larger reprint of this scan can be found in Figure 11.6 on page 103.

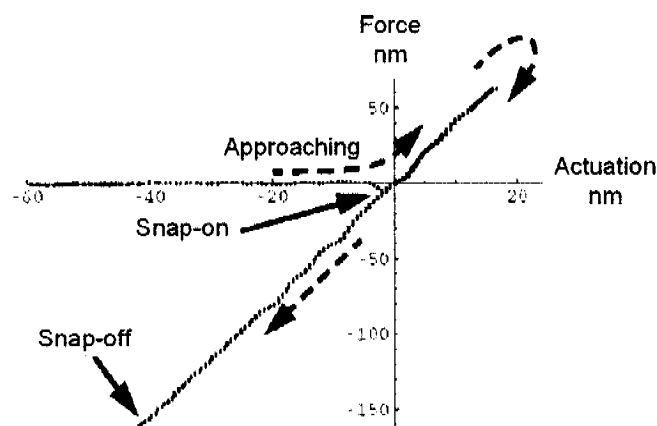


Figure 10.10: Typical force-distance plot describing the phases of a measurement loop. The dotted arrows point out the trajectory.

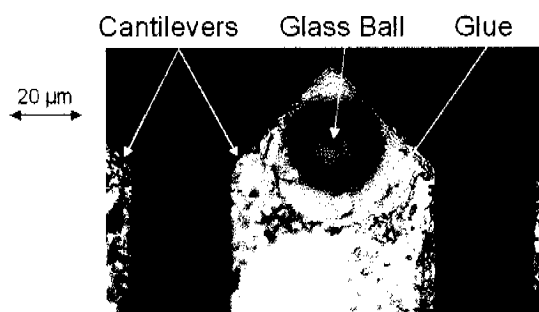


Figure 10.11: Spherical glass beads ($20\text{ }\mu\text{m}$ diameter) affixed to the cantilever end for force-distance measurements. The contact radius is approximately 100 nm and the contact area is $0.03\text{ }\mu\text{m}^2$ [Hafizovic et al., 2004].

10.4.2 Force Distance Measurements

In the force-distance mode, the microsystem was operated with the on-chip controllers performing the approach to and retraction from the sample surface, during which the force on the cantilever was recorded. Figure 10.10 shows a typical force-distance plot describing the phases of a measurement loop. The dotted arrows point out the trajectory.

A ramp signal commands the thermal actuator until the tip reaches the sample (“snap-on”). Then, the tip is retracted by another ramp signal until the tip is released from the sample (“snap-off”). The cantilever deflection (tip-surface normal force) is recorded to calculate the interaction force between the tip and sample using the height of the hysteresis at the “snap-off” point and the spring constant of the cantilever. Each force-distance measurement (complete trajectory) was conducted in 0.5 seconds. The experimentally determined force resolution was better than 1 nN . To achieve a welldefined geometry of the cantilever surface to interact with the sample, spherical glass beads ($20\text{ }\mu\text{m}$ diameter) were attached to the cantilever end (Figure 10.11). The radius of the contact spot is approx-

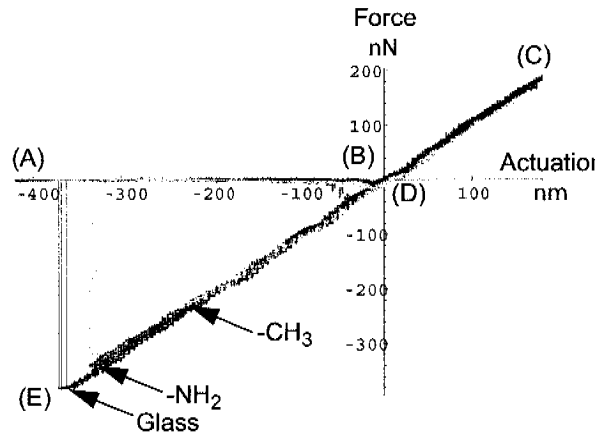


Figure 10.12: Four consecutive force response measurements of a glass surface and of gold surfaces that were coated with a methylterminated ($-\text{CH}_3$) and an amino-terminated ($-\text{NH}_2$) undecanethiol self-assembled monolayer (SAM). The cantilever slowly approaches the surface (A) until the bead experiences an attractive force and comes in contact with the surface (“snap on”, (B)). The load on the cantilever is increased up to about 160 nN (C), then, the cantilever actuation is reversed. At (D) the force exerted on the sample changes from compressive to tensile, and at (E) the force is large enough to pull the bead off the surface (“snap off”) [Hafizovic et al., 2004].

imately 100 nm and the contact area has been calculated to $0.03 \mu\text{m}^2$ [Franks et al., 2002]. Three identical glass beads have been affixed to three cantilevers of the same chip for surface probing. The samples consisted of glass chips with gold patterns that were coated either with a methyl-terminated ($-\text{CH}_3$) or with an amino-terminated ($-\text{NH}_2$) undecanethiol self-assembled monolayer (SAM). Figure 10.12 shows the corresponding force response measurements. Four consecutive measurements for each sample surface are shown. Each data point represents the average of twelve recorded values. The data have been low-pass filtered with a cut-off frequency of 30 Hz. It is evident from Figure 10.11 that the glass bead (polar surface) interacts most intensely with the glass sample surface, to a lesser extent with the amino-terminated SAM, and the weakest interaction is between the nonpolar methyl-terminated SAM and the glass sphere.

10.5 Conclusions

The CMOS monolithic AFM microsystem reported in this paper is an autonomous unit that allows for precise and fully-controlled mechanical manipulation at nanometer / nanonewton resolution. The microsystem was fabricated using a standard $0.8 \mu\text{m}$ CMOS process with some post-CMOS micromachining steps. The microsystem characterization evidenced a vertical resolution of better than 1 nm and a force resolution of better than 1 nN. Table 10.3 summarizes the performance and specification of the CMOS monolithic AFM microsystem.

Table 10.3: Performance and specification summary of the CMOS monolithic AFM microsystem.

Parameter	Value
Maximum z-range	
High resolution	0.5 μm
Low resolution	6 μm
Bandwidth	3 kHz
Maximum offset voltage	100 mV
Total power consumption	140 mW @ 5 V
Clock frequency	1 MHz
Technology	0.8 μm CMOS
Die area	70 mm ²

For operating the microsystem, only a simple printed circuit board is necessary to provide some reference voltages.

The CMOS monolithic AFM microsystem paves the way to developing other smart microsystems that will further bridge the gap to the nanoworld.

Acknowledgments

The authors thank Prof. Henry Baltes (on leave) for sharing laboratory resources and for his ongoing stimulating interest in their work. The authors would also like to thank Dr. Stefano Taschini for his help in the digital design and Wendy Franks and Frauke Greve for providing the neurite and self-assembled-monolayer sample surfaces.

Chapter 11

Single-Chip Mechatronic Microsystem for Surface Imaging and Force Response Studies

S. Hafizovic, D. Barrettino, T. Volden, J. Sedivy, K.-U. Kirstein,
O. Brand, and A. Hierlemann

Proc Natl Acad Sci USA. 2004 Dec 7, 101(49), pp. 17011-17015

Physical Electronics Laboratory, ETH Zurich, Switzerland

Abstract - We report on a stand-alone single-chip ($7 \times 10 \text{ mm}^2$) atomic force microscopy unit including a fully integrated array of cantilevers, each of which has an individual actuation, detection, and control unit so that standard atomic force microscopy operations can be performed by means of the chip only without any external controller. The system offers drastically reduced overall size and costs as well as increased scanning speed and can be fabricated with standard complementary metal oxide semiconductor technology with some subsequent micromachining steps to form the cantilevers. Full integration of microelectronic and micromechanical components on the same chip allows for the controlling and monitoring of all system functions. The on-chip circuitry, which includes analog signal amplification and filtering stages with offset compensation, analog-to-digital converters, a powerful digital signal processor, and an on-chip digital interface for data transmission, notably improves the overall system performance. The microsystem characterization evidenced a vertical resolution of $<1 \text{ nm}$ and a force resolution of $<1 \text{ nN}$ as shown in the measurement results. The monolithic system represents a paradigm of a mechatronic microsystem that allows for precise and fully controlled mechanical manipulation in the nanoworld.

11.1 Introduction

New measurement, metrology and imaging techniques have been pivotal to the rapid development of many branches of science such as materials science, microelectronics and microbiology to name a few. The invention of the scanning-tunneling microscope (STM) by Binnig and Rohrer in 1982 [Binnig and Rohrer, 1982] and, in particular, the invention of the atomic-force microscope (AFM) in 1986 [Binnig et al., 1986] have established a basis for many findings in various scientific areas over the last years. The AFM has evolved at an exceptional speed from a laboratory prototype to a commercial instrument (see, e.g., Veeco Instruments, Woodbury, NY) and many other scanning probe techniques [Meyer and Heinzelmann, 1995] have been developed, which will not be further mentioned here. The AFM has been used to measure forces during stretching of DNA strands [Rief et al., 1999] and rupture forces of single covalent bonds [Grandbois et al., 1999] and can be operated in liquids, which enables its use in biological applications [Lehenkari et al., 2000, You et al., 2000, Bowen et al., 1998]. The AFM also can be used to perform surface manipulations such as lithographic fabrication of a transistor [Minne et al., 1999, 1996], or AFM-based storage [Despont et al., 2000, Brugger et al., 1992].

Commercially available AFM instruments are rather bulky and have a low throughput due to the serial nature of the involved scanning process and the limited scanning range, which renders the investigation of larger samples rather laborious. The detection of the cantilever deflection is done mostly by means of a laser, which is costly and makes the adjustment and cantilever exchange very time-consuming, in particular, when operating in vacuum environment. To overcome these limitations, AFM probes with integrated detection schemes such as capacitive [Brugger et al., 1992], piezoelectric, or piezoresistive schemes [Minne et al., 1999, Linnemann et al., 1995, Gotszalk et al., 1998, Jumpertz et al., 1998, Thaysen et al., 2000], and high-speed scanning systems that rely on arrays of cantilevers featuring piezoelectric excitation and piezoresistive/piezoelectric readout [Minne et al., 1999, 1998b,a, Y.-S.Kim et al., 2003] have been developed. All of those systems, however, require a larger set of desktop equipment since no integrated electronics or functions are provided.

11.2 Materials and Methods

We report here on a standalone single-chip AFM unit including a fully integrated array of cantilevers, each of which has its individual actuation, detection, control, amplification and on-chip digital processing unit as well as its individual offset compensation so that, e.g., constant-force operation is performed without any external controller. The cantilevers can be moved and precisely controlled within a range of 0.5 to 6 μm (at 0.5 to 6 nm resolution) so that the chip has only to be brought within a distance of maximal 6 μm to a surface and then

can be used to carry out, e.g., multiple force-distance measurements using the integrated electronics via a Labview interface. Only a x-y-scanning stage is necessary for imaging or lateral scanning by using one or several of the individually controlled cantilevers. The individual cantilever control is of paramount importance in using multicantilever arrays since there is always a slight tilt between the chip plane determining the cantilever positions and the sample surface plane so that some cantilevers are closer to the surface than others. Parallel scanning of multiple cantilevers only provides good results by implementing a fast and simultaneous individual force feedback for each scanning cantilever in a closed loop (no multiplexing).

Earlier work on integrated AFM instruments included a chip featuring up to 10 cantilevers and analog time-multiplexed actuation and readout of the cantilevers featuring rather limited force and vertical resolution [Lange et al., 2002a].

The novel fully integrated AFM microsystem is fabricated in standard integrated-circuit technology (complementary metal oxide semiconductor, CMOS). The cantilevers of this microsystem are deflected by making use of the bimorph effect, i.e., the different thermal expansion coefficients of silicon and the aluminum layers (metal 1 and metal 2 of the CMOS process, see refs. [Lange et al., 2002a, Akiyama et al., 2002]). Upon heating the layer sandwich, the cantilever bends, and the bending can be precisely controlled through the heating current passing the diffused heating resistor at the surface of the silicon. To detect the cantilever deflection upon force exerted on the tip, four piezoresistors (resistors that change their resistance upon mechanical stress) are arranged at the cantilever base in a Wheatstone-bridge configuration [Gotszalk et al., 1998, Jumpertz et al., 1998, Lange et al., 2002a]. Two of them are collaterally oriented to the cantilever axis, two of them are oriented perpendicularly in order to achieve a maximum signal upon deflection (see also Figure 11.1) [Lange et al., 2002a]. A detailed 3-dimensional schematic of the scanning microcantilever is shown in Figure 11.2. The cantilever is 500 μm long, 85 μm wide, 5 μm thick and features a force constant of 1 N/m. The integrated circuitry is designed to perform the microcantilever deflection control in vertical direction, an offset compensation, and the amplification and conditioning of the signals from the piezoresistive Wheatstone bridge. The details of this integrated circuitry are described later. The vertical drift of the individually controlled cantilever was measured to be on the order of 10 nm per hour under ambient conditions.

Functional elements of the cantilevers such as the heaters or the metal components are fabricated during the CMOS processing (0.8- μm double-metal, double-poly CMOS process of Austriamicrosystems, Austria). The cantilever structures are then formed in subsequent micromachining steps. First, anisotropic silicon wet-etching using a potassium-hydroxide solution applied to the backside of the wafer is used to create silicon membranes. An electrochemical etch stop technique provides a defined and uniform thickness of the membrane, which consists of the n-well layer of the CMOS process [Müller et al., 2000]. Front-side reac-

Figure 11.1: Close-up of the cantilever area. The chip features twelve cantilevers, ten of which can be potentially used for scanning (500 μm long, 85 μm wide). Only the four cantilevers in the center are connected to the circuitry. The two shorter ones at the flanks (250 μm long) serve as references.

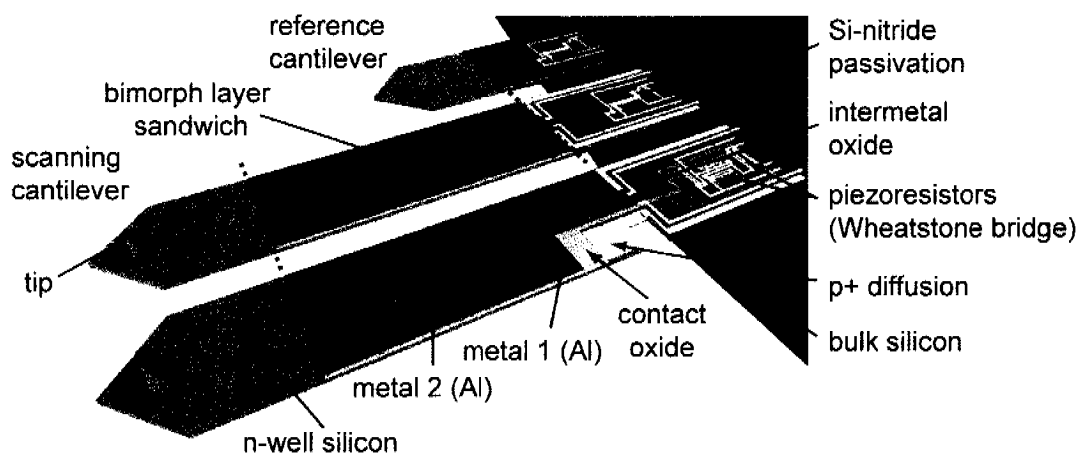
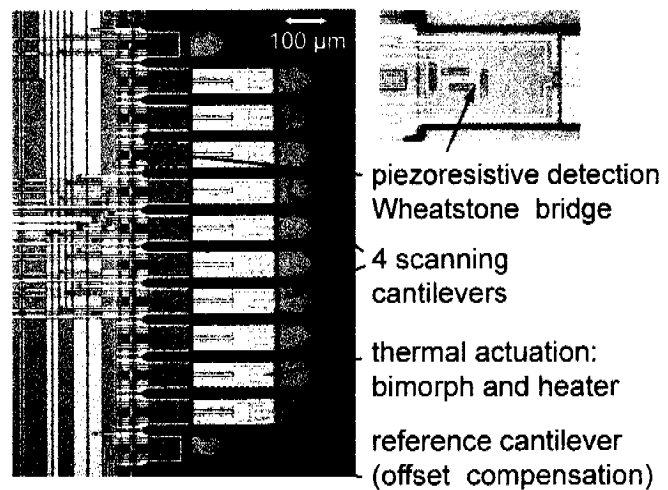


Figure 11.2: Three-dimensional schematic of three cantilevers (two scanning cantilevers and one shorter reference cantilever) showing the different CMOS layers as they are used for the various cantilever components (actuation: bimorph and heater; detection: piezoresistive Wheatstone bridge).

digital block including digital
signal processor (DSP), 4 serial
interfaces, 4 digital controllers,
4 filters (sensor-actuator coupling)

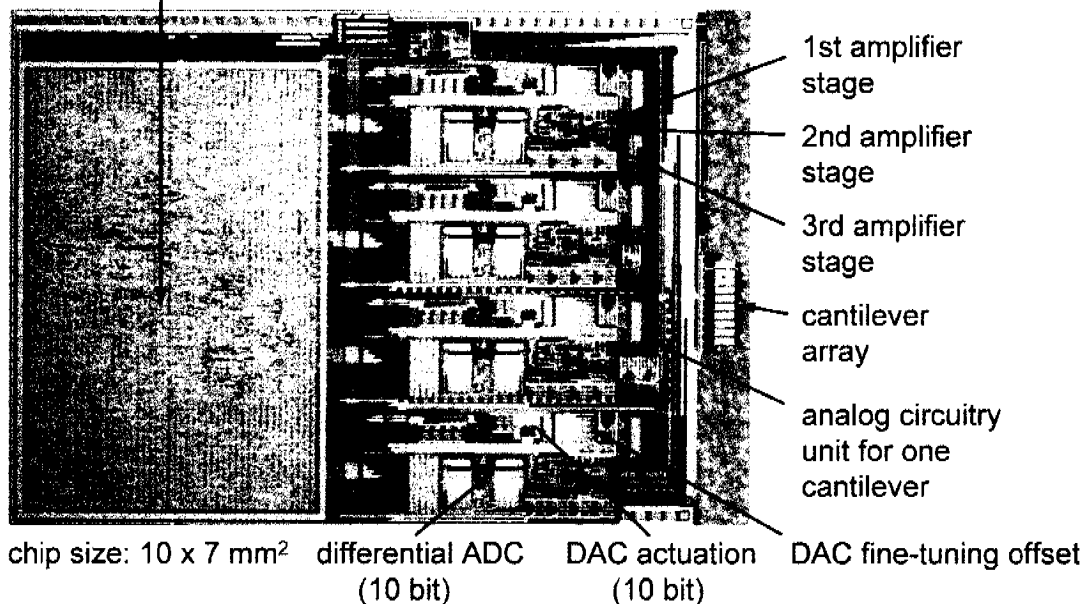


Figure 11.3: Micrograph of the overall microsystem chip featuring the digital block on the left side and four identical analog units to control the four central cantilevers on the right side. The different circuitry subunits are indicated (ADC: analog-to-digital converter; DAC: digital-to-analog converter).

tive ion etching (RIE) or wet etching is used afterwards to locally remove oxide layers from the cantilever tip, then front-side RIE is used again to release the cantilevers. Finally, a silicon nitride tip (radius 10-20 nm) is mounted at the cantilever end.

Figure 11.3 shows a micrograph of the overall microsystem chip, which features a die size of 10.00 mm \times 7.00 mm, and Figure 11.1 shows a close-up of the 12-cantilever array. The spacing between the cantilevers is 25 μ m. The ten scanning cantilevers, four of which are connected to the circuitry and can be individually controlled (Figure 11.1), are located in the center of the array. The two reference cantilevers located at the left and right end of the array are 250 μ m long, 85 μ m wide and 5 μ m thick. They feature only the Wheatstone bridge for deflection detection, and they are shorter than the scanning cantilevers in order not to contact the sample surface. The reference cantilevers are used for offset compensation of the Wheatstone bridge.

The electronics covering most of the chip area include four repeated mainly analog circuitry units (Figures 11.3 and 11.4), which are connected to the four central cantilevers and a common digital block, which is used for digital signal processing and includes a serial digital bus interface to connect to off-chip components. Due

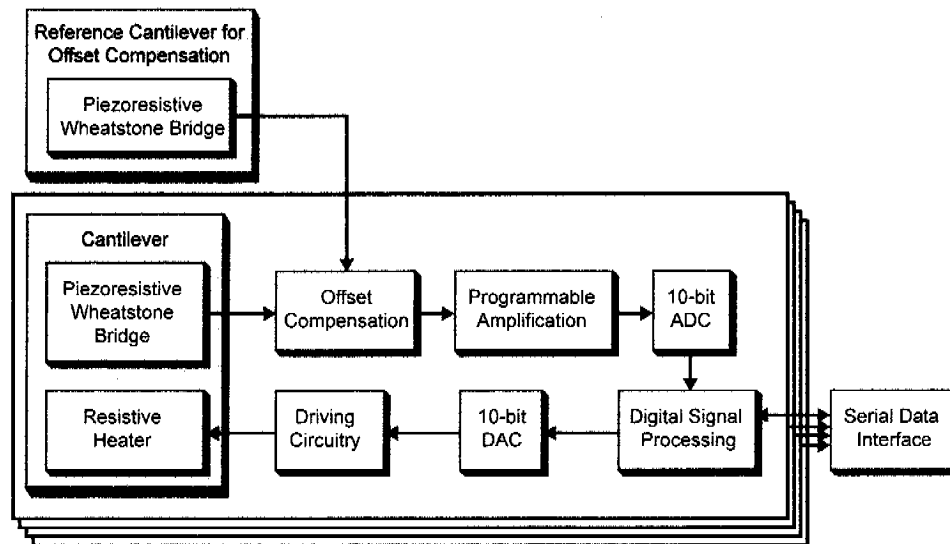


Figure 11.4: Schematic of the chip architecture and the electronic components. The four frames indicate that these subunits are repeated for each active cantilever, i.e., four times. The chip also includes a digital signal processor and serial digital interface (ADC: analog-to-digital converter; DAC: digital-to-analog converter).

to chip area constraints, only four readout and digital processing channels have been realized on-chip to provide a proof of concept. The integration of 10 readout-channels can be realized by a redesign in CMOS technology with smaller feature size.

A simplified block diagram of the major circuitry units is shown in Figure 11.4. The microsystem has an analog-digital mixed-signal architecture. The cantilever deflection signal coming from the Wheatstone bridge is amplified, filtered, conditioned and translated into the digital domain. All control operations are handled in the digital signal processor, which then issues actuation signals to the cantilever heater via a digital-to-analog converter. This way, a closed-loop operation and self-controlled system is realized for every cantilever.

The circuitry unit repeated with each cantilever includes a fully differential analog amplification (three-stage amplification: fully differential low-noise programmable amplifier, programmable amplifier, fixed amplifier) and filtering unit, which provides low-noise readout and signal processing of the Wheatstone bridge signal. Multiple stages of offset compensation ensure the maximum possible force resolution to be achieved: After the first amplifier stage the coarse offset of the piezoresistive Wheatstone bridge is compensated by subtracting the offset signal of the reference cantilever structure (Figure 11.1). This offset is mainly due to fabrication spread in the cantilever and the Wheatstone bridge [Volden et al., 2004] and also includes the offsets of the signal amplifiers. The imperfect matching of the diffused piezoresistive sensors from cantilever to cantilever can be compensated by another compensation stage, which enables an individual ad-

justment of the DC value of each readout channel with an 8-bit resolution by means of a programmable digital-to-analog converter (DAC). The overall gain of the analog signal processing is programmable from 18 dB to up to 44 dB to cover the whole range of expectable forces in various applications. The amplified force signal is converted by a 10-bit successive-approximation analog-to-digital converter (ADC) and fed into the digital signal-processor (DSP). The DSP unit comprises two programmable infinite-impulse-response (IIR) filters, each with 6 coefficients, which can be configured to act as proportional-integral-derivative (PID) controllers for the constant-force imaging mode. The controllers also provide averaging functions to further improve the force resolution and they compensate for thermal sensor-actuator crosstalk. The computing power of the DSP unit for the four cantilevers amounts to 16 million arithmetic operations per second, which is one of the highest values ever realized in CMOS-based micromechanical systems. This large value enables to issue 100,000 actuation signals per second per cantilever for repositioning and allows for precise cantilever position control even when fast force changes are to be expected. The cantilever actuation signals as coming from the DSP unit are converted to the analog domain by 10-bit flash DACs and provide together with analog square-root circuits linear actuation characteristics. For a total deflection range of 1 μm , a resolution of 1 nm has been achieved, which can be further improved to 0.5 nm by lowering the overall deflection range to 0.5 μm .

11.3 Results and Discussion

Two prototype applications were selected to show the performance of the monolithic AFM microsystem, (a) surface imaging and (b) force distance measurements.

For surface imaging, the x-y scanning function of a Nanoscope III, Digital Instruments, USA was used, and the microsystem was operated in contact mode. The on-chip force controller measures the force acting on the cantilever and keeps it constant while the tip is scanned over the surface. Height information of the scanned sample is obtained from the actuation signal that is required to keep the cantilever force constant. Figure 11.5 shows a scanning image of a silicon grating with 18 nm steps at 3 μm distance recorded at a scanning speed of 20 $\mu\text{m/s}$ and a force of 50 nN. Figure 11.5.b shows a small-range line scan recorded at 3 $\mu\text{m/s}$. The maximum possible scanning speed of the monolithic system is approximately 1 mm/s. The maximum readout rate of the force signal is 100 kHz, which allows for averaging multiple values per data point in the DSP for better noise suppression. Each displayed data point in Figure 11.5 represents the average of 300 values. The measured error signal indicates an excellent tracking of the surface topography. A vertical resolution of better than 1 nm has been achieved (Figure 11.5.b). A larger area-scan of a biological sample at a force of 10 nN and a scanning speed of 100 $\mu\text{m/s}$ is shown in Figure 11.6 representing a

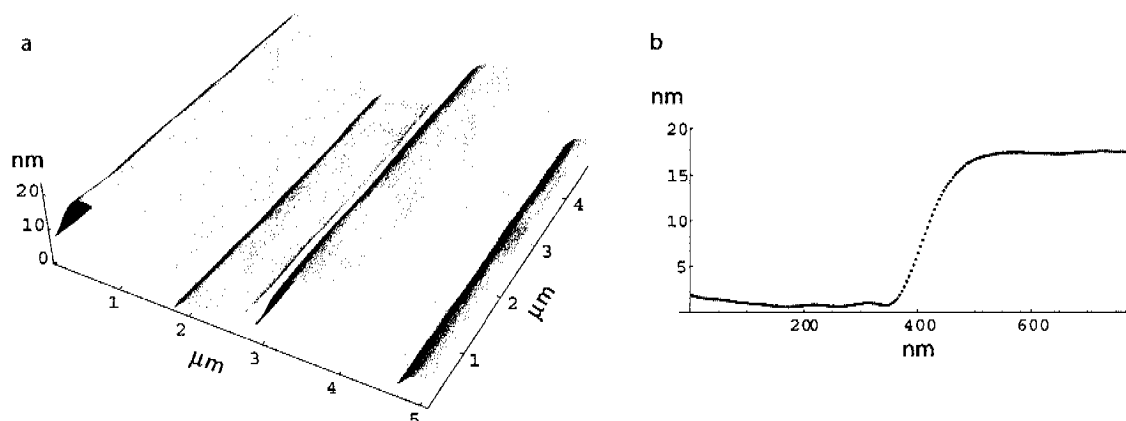


Figure 11.5: (a): Scanning image (constant-force mode) of a silicon grating featuring step heights of 18 nm at 3 μm distance. A Gaussian smoothing with a radius of 3 data points was applied to yield the image.

(b): High-resolution line-scan over one 18-nm step. Each displayed data point represents the average of 300 values. The vertical resolution is better than 1 nm.

network of dried-out chicken neurites on a silicon oxide surface after fixation. A light-microscope image is given for comparison (Figure 11.7). The force exerted on a sample can be kept as low as 5 nN so that also soft samples can be imaged. Tapping-mode operation can be also realized with the monolithic AFM system but was not tested so far.

In the force-distance mode, the microsystem was operated with the on-chip controllers performing the approach to and retraction from the sample surface, during which the force on the cantilever was recorded. Two force-distance measurements were conducted per second. The experimentally determined force resolution was better than 1 nN.

To achieve a well-defined geometry of the cantilever surface to interact with the sample, spherical glass beads (20 μm diameter) were attached to the cantilever end (Figure 11.8). The radius of the contact spot is approximately 100 nm and the contact area has been calculated to 0.03 μm^2 [Franks et al., 2002]. Three identical glass beads have been affixed to three cantilevers of the same chip for surface probing. The samples consisted of glass chips with gold patterns that were coated either with a methyl-terminated or an amino-terminated undecanethiol self-assembled monolayer. Figure 11.9 shows the corresponding force response measurements. Four consecutive measurements for each sample surface are shown. Each data point represents the average of twelve recorded values. The data have been low-pass filtered with a cut-off at 30 Hz corner frequency. It is evident from Figure 11.10, that the glass bead (polar surface) interacts most intensely with the glass sample surface, to a lesser extent with the amino-terminated self-assembled monolayer (SAM) and the weakest interaction is between the nonpolar methyl-terminated SAM and the glass sphere. The reproducibility of the measurements is shown in Figure 11.10. The “snap-on” and

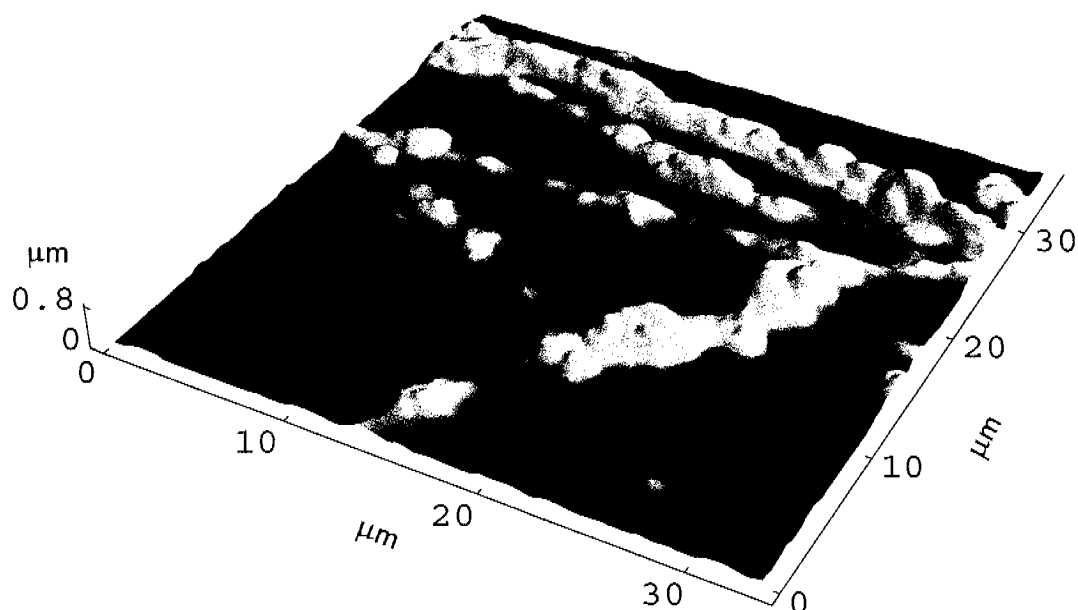


Figure 11.6: Scan of a biological sample at a force of 10 nN and a scanning speed of 100 $\mu\text{m/s}$ representing a network of dried-out chicken neurites on a silicon oxide surface after fixation.

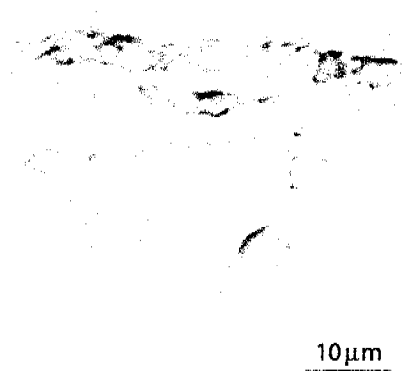


Figure 11.7: Light-microscope image of the same area of a network of dried-out chicken neurites for comparison.

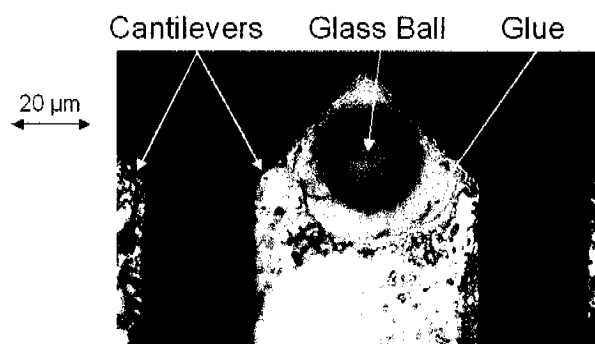


Figure 11.8: Spherical glass beads (20 μm diameter) affixed to the cantilever end for force-distance measurements. The contact radius is approximately 100 nm and the contact area is 0.03 μm^2 .

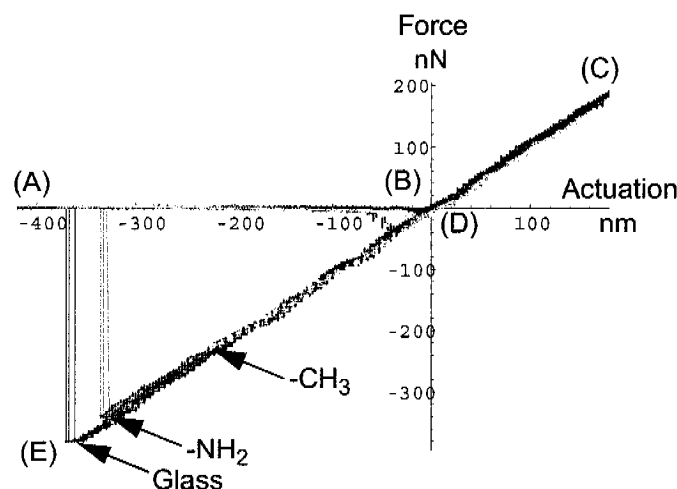


Figure 11.9: Left: Four consecutive force response measurements of a glass surface and of gold surfaces that were coated with a methyl-terminated and an amino-terminated undecanethiol self-assembled monolayer. The cantilever slowly approaches the surface (A) until the bead experiences an attractive force and comes in contact with the surface (“snap-on”, (B)). The load on the cantilever is increased up to about 160 nN (C), then, the cantilever actuation is reversed. At (D) the force exerted on the sample changes from compressive to tensile, and at (E) the force is large enough to pull the bead off the surface (“snap-off”).

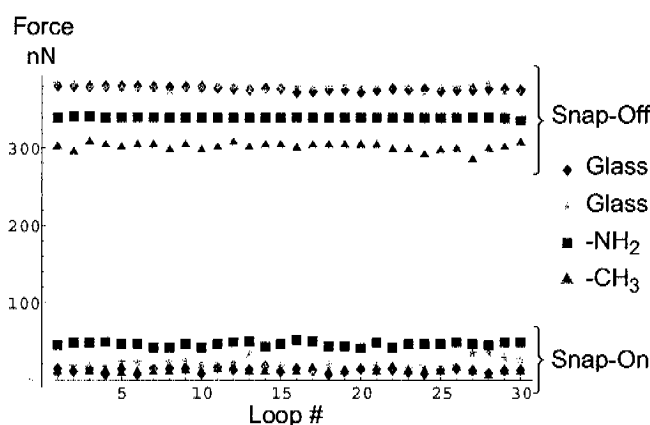


Figure 11.10: “Snap-on” and “snap-off” forces of 30 consecutive measurements (same cantilever). The mean values and standard deviations of the “snap-off” forces amount to 379 nN (2.6 nN) for glass, to 340 nN (0.9 nN) for the amino-terminated SAM, and to 301 nN (4.9 nN) for the methyl-terminated SAM.

“snap-off” forces of 30 consecutive measurements (same cantilever) are displayed. The mean values and standard deviations of the “snap-off” forces amount to 379 nN (2.6 nN) for glass, to 340 nN (0.9 nN) for the amino-terminated SAM, and to 301 nN (4.9 nN) for the methyl-terminated SAM. The same measurements conducted with the other two cantilevers of the same chip showed some fluctuations in the absolute values, but the relations between glass and amino-terminated SAM as well as that between glass and methyl-terminated SAM were preserved within 5% to 20% relative error.

For operating the monolithic AFM-microsystem, only a simple circuit board is necessary to provide power supply stabilization and some reference voltages. There is no need for external signal processing capacity either for closed-loop imaging or open-loop force measurements. All operations can be performed on chip. The digital interface connects the system to a computer for data capturing and visualization.

In summary, the developed AFM-microsystem is a monolithic autonomous unit that allows for precise and fully controlled mechanical manipulation at nanometer/nanonewton resolution. The device is applicable to many fields including aerospace (low payload and low power), biotechnology (cell manipulation, force detection) and security (microbalance operation) applications. In particular the force-distance mode can be used without scanning stage so that surface and material characterization using, e.g., differently modified cantilevers are easy to perform. The chip can also be immersed in liquid phase. The device paves the way to developing other smart microsystems that will further bridge the gap to the nanoworld.

Acknowledgments

The authors thank Prof. Henry Baltes (on leave) for sharing laboratory resources and for his ongoing stimulating interest in their work. The authors also are indebted to former coworkers, notably Dirk Lange and Christoph Hagleitner, for developing a pre-version of this fully integrated AFM chip. Wendy Franks and Frauke Greve are acknowledged for providing the SAM-modified surface and the neuron sample.

Chapter 12

Outlook

Future integrated AFMs will feature different floor plans with different area fractions covered by the transducer, the analog and the digital circuitry: The cantilever array will occupy only a small fraction of the chip. Analog circuits can be compacted since more a-priori knowledge of the cantilever characteristics, such as offset etc. is available. Analog-to-digital conversion is likely to be more complex to provide larger bandwidths. The largest part will be the digital core. The focus of future development work will be on increasing the system precision, on providing larger bandwidths, on integrating digital phase-lock loops and on further reducing the number of interconnections.

New actuation schemes such as magnetic actuation (see [Lange et al., 2002b]) have to be developed that can be applied in also a liquid-phase environment to satisfy the needs of the rapidly expanding field of bioanalytics. The packaging of an integrated AFM system for operation in liquid-phase environments will pose further challenges.

Chapter 13

AFM Bibliography

- T. Akiyama, U. Staufer, and N. F. de Rooij. Fast driving technique for integrated thermal bimorph actuator toward high-throughput atomic-force microscopy, review of scientific instruments. *Review of Scientific Instruments*, 73(7):2643–2646, 2002.
- AMS. Austriamicrosystems. <http://www.austriamicrosystems.com>.
- R.J. Baker, H.W. Li, and D.E. Boyce. *CMOS Circuit Design, Layout and Simulation*. IEEE Press, USA, 1998.
- M. Banu, J. Khoury, and Y. Tsvividis. Fully differential operational amplifiers with accurate output balancing. *IEEE Journal of Solid-State Circuits*, 23:1410–1414, 1988.
- D. Barrettino. *CMOS Readout and Control Architectures for Monolithic Hotplate and Cantilever Microsystems*, Ph.D. thesis Nr. 15412. ETH Zurich, Zurich, 2004.
- D. Barrettino, M. Graf, S. Taschini, M. Zimmermann, C. Hagleitner, A. Hierlemann, and H. Baltes. Hotplate-based conductometric monolithic cmos gas sensor system. In *IEEE Symposium on VLSI Circuits*, pages 157–160, Kyoto, Japan, 2003.
- G. Binnig and H. Rohrer. Scanning tunneling microscopy. *Helv. Phys. Acta*, 55:355, 1982.
- G. Binnig, C.F. Quate, and Ch. Gerber. Atomic force microscopy. *Phys. Rev. Lett.*, 56:930, 1986.
- W. R. Bowen, N. Hilal, R. W. Lovitt, and C. J. Wright. Direct measurement of the force of adhesion of a single biological cell using an atomic force microscope. *Colloids Surf. A*, 136:231–234, 1998.
- J. Brugger, R.A. Buser, and N.F. de Rooij. Micromachined atomic force microprobe with integrated capacitive read-out. *Journal of Micromechanics and Microengineering*, 2:218–224, 1992.

- M. Despont, J. Brugger, U. Drechsler, U. Durig, W. Haberle, M. Lutwyche, H. Rothuizen, R. Stutz, R. Widmer, G. Binnig, H. Rohrer, and P. Vettiger. Vlsi-nems chip for parallel afm data storage. *Sens. Actuators A*, 80(2):100–107, 2000.
- W. Franks, D. Lange, S. Lee, A. Hierlemann, N. Spencer, and H. Baltes. Nanochemical surface analyzer in cmos technology. *Ultramicroscopy*, 91:21–27, 2002.
- T. Gotszalk, J. Radojewski, P.B. Grabić, P. Dumania, F. Shi, P. Hudek, and I.W. Rangelow. Fabrication of multipurpose piezoresistive wheatstone bridge cantilevers with conductive microtips for electrostatic and scanning capacitance microscopy. *Journal of Vac. Sci. Technol. B*, 16:3948–3953, 1998.
- M. Grandbois, M. Beyer, M. Rief, H. Claussen-Schaumann, and H. Gaub. How strong is a covalent bond. *Science*, 283:1727–1730, 1999.
- S. Hafizovic, D. Barrettino, T. Volden, J. Sedivy, K. Kirstein, O. Brand, and A. Hierlemann. Single-chip mechatronic microsystem for surface imaging and force response studies. *Proc Natl Acad Sci USA.*, 101(49):17011–17015, 2004.
- S. Hafizovic, K.-U. Kirstein, and A. Hierlemann. *Integrated Cantilevers and Atomic Force Microscopes*, volume 5, chapter 1. Springer, 2007.
- Veeco Instruments Inc. <http://www.veeco.com/>.
- R. Jumpertz, A. v. d. Hart, O. Ohlsson, F. Saurenbach, and J. Schelten. Piezoresistive sensors on afm cantilevers with atomic resolution. *Microelectronic Engineering*, 41/42:441–444, 1998.
- Y.S. Kim, H.J. Nam, S.M. Cho, J.W. Hong, D.C. Kim, and J. U. Bu. Pzt cantilever array integrated with piezoresistor sensor for high speed parallel operation of afm. *Sens. Actuators A*, 103:122–129, 1998.
- K. Laker and W. Sansen. *Design of Analog Integrated Circuits and Systems*. McGraw-Hill Professional, Singapore, 1994.
- D. Lange, O. Brand, and H. Baltes. *CMOS Cantilever Sensor Systems*. Springer-Verlag, Berlin, 2002a.
- D. Lange, C. Hagleitner, C. Herzog, O. Brand, and H. Baltes. Electromagnetic actuation and mos-transistor sensing for cmos-integrated micromechanical resonators. *Sens. Actuators A*, 6:150–155, 2002b.
- P. P. Lehenkari, G. T. Charras, A. Nykänen, and M. A. Horton. Adapting atomic force microscopy for cell biology. *Ultramicroscopy*, 82:289–295, 2000.
- R. Linnemann, T. Gotszalk, L. Hadjiiski, and I.W. Rangelow. Characterization of a cantilever with an integrated deflection sensor. *Thin Solid Films*, 264:159–164, 1995.
- H.J. Mamin, R. Ried, B. Terris, and D. Rugar. High-density data storage based on the atomic force microscope. *Proceedings of the IEEE*, 87:1014–1027, 1999.

- E. Meyer and H. Heinzelmann. *Scanning Force Microscopy in Scanning Tunneling Microscopy II*, eds. Wiesendanger and R., Güntherodt and H.J. Springer, Heidelberg, 1995.
- S.C. Minne, S.R. Manalis, A. Atalar, and C.F. Quate. Independent parallel lithography using the atomic force microscope. *J. Vac. Sci. Technol. B*, 14:2456–2461, 1996.
- S.C. Minne, J.D. Adams, G. Yaralioglu, S. R. Manalis, A. Atalar, and C.F. Quate. Centimeter scale atomic force microscope imaging and lithography. *Applied Physics Letters*, 73:1742–1743, 1998a.
- S.C. Minne, G. Yaralioglu, S. R. Manalis, J.D. Adams, J. Zesch, A. Atalar, and C.F. Quate. Automated parallel high-speed atomic force microscopy. *Applied Physics Letters*, 72:2340–2342, 1998b.
- S.C. Minne, S.R. Manalis, and C.F. Quate. *Bringing Scanning Probe Microscopy up to Speed*. Kluwer Academic Publishers, Boston, 1999.
- T. Müller, M. Brandl, O. Brand, and H. Baltes. An industrial cmos process family adapted for the fabrication of smart silicon sensors. *Sensors and Actuators A*, 84: 126–133, 2000.
- M. Rief, H. Claussen-Schaumann, and H. Gaub. Sequence-dependent mechanics of single dna molecules. *Nature Structural Biology*, 6:346–349, 1999.
- J. J. F. Rijns. Cmos low-distortion high-frequency variable-gain amplifier. *IEEE Journal of Solid-State Circuits*, 31(7):1029–1034, 1996.
- J. Thaysen, A. Boisen, O. Hansen, and S. Bouwstra. Atomic force microscopy probe with piezoresistive read-out and a highly symmetrical wheatstone bridge arrangement. *Sensors and Actuators A*, 83:47–53, 2000.
- P. Vettiger, G. Cross, M. Despont, U. Drechsler, U. Durig, B. Gotsmann, W. Haberle, M.A. Lantz, H.E. Rothuizen, R. Stutz, and G.K. Binnig. The "millipede" - nanotechnology entering data storage. *IEEE Transactions on Nanotechnology*, 1:39–55, 2002.
- T. Volden, M. Zimmermann, D. Lange, O. Brand, and H. Baltes. Dynamics of cmos-based thermally actuated cantilever arrays for force microscopy. *Sensors and Actuators A*, 115(2-3):516–522, 2004.
- Y.-S. Kim, H.-J. Nam, S.-M. Cho, J.-W. Hong, D.-C. Kim, and Jong U. Bu. Pzt cantilever array integrated with piezoresistor sensor for high speed parallel operation of afm. *Sensors and Actuators A*, 103(1-2):122–129, 2003.
- H.X. You, J.M. Lau, S. Zhang, and L. Yu. Atomic force microscopy imaging of living cells: A preliminary study of the disruptive effect of the cantilever tip on cell morphology. *Ultramicroscopy*, 82:297–305, 2000.

Appendix A

Glossary

Abbreviations

ADC	analog-to-digital converter
ASIC	application-specific integrated circuits
AFM	atomic force microscope
DSP	digital signal processor
FPGA	field programmable gate array. Reprogrammable logic device
IC	integrated circuit
IIR	infinite-input-response filter. Filter with feedback.
CMOS	complementary metal oxide semiconductor
DAC	digital-to-analog converter
LSM	liquid state machine
MAC	multiplier accumulator
MEA	multielectrode array
MEMS	microelectromechanical systems
NN	neuronal networks
NNN	natural neuronal network
PCB	printed circuit board
PID	proportional-integral-derivative (controller)

RMS	root mean square $\sqrt{1/n \sum^n x_i^2}$
SNR	signal-to-noise ratio
SAC	sensor-actuator crosstalk
ULSI	ultra large-scale integration
VLSI	very large-scale integration

Appendix B

Publications

Journal Papers

1. *A CMOS-based Microelectrode Array for Interaction with Neuronal Cultures*
S. Hafizovic, F. Heer, T. Ugniwenko, A. Blau, C. Ziegler and A. Hierlemann, submitted to Journal of Neuroscience Methods
2. *Single-chip microelectronic system to interface with living cells*
F. Heer, **S. Hafizovic**, T. Ugniwenko, W. Franks, A. Blau, C. Ziegler, J.-C. Perriard and A. Hierlemann, in press, Biosensors & Bioelectronics
3. *CMOS microelectrode array for bidirectional interaction with neuronal networks*
F. Heer, **S. Hafizovic**, W. Franks, A. Blau, C. Ziegler and A. Hierlemann, IEEE Journal of Solid-State Circuits, 2006; 41: 1620-1628
4. *Validation of X-ray lithography and development simulation system for moving mask deep X-ray lithography*
Y. Hirai, **S. Hafizovic**, N. Matsuzuka, J.G. Korvink and O. Tabata, JMEMS 2006, 15(1), pp. 159-168
5. *CMOS monolithic metal-oxide gas sensor microsystems*
D. Barrettino, M. Graf, S. Taschini, **S. Hafizovic**, C. Hagleitner, A. Hierlemann, IEEE Sensors Journal 2006, 6(2), pp. 276- 286
6. *Characterization of a microfluidic dispensing system for localised stimulation of cellular networks*
T. Kraus, E. Verpoorte, V. Linder, W. Franks, A. Hierlemann, F. Heer, **S. Hafizovic**, T. Fujii, N.-F.-de Rooij and S. Koster, Lab on a Chip 2006, 6, pp. 218-229
7. *CMOS monolithic mechatronic microsystem for surface imaging and force response studies*

- D. Barrettino, **S. Hafizovic**, T. Volden, J. Sedivy, K.-U. Kirstein and A. Hierlemann, *Journal of Solid-State Circuits (JSSC)* 2005, 40(4), pp. 951-959
8. *CMOS biosensor with guided cell growth*
W. Franks, F. Heer, S. Tosatti, **S. Hafizovic**, P. Seif, M. Textor and A. Hierlemann, *Proc. IEEE Sensors*, 2004, vol. 2, pp. 1014-1017
 9. *Single-chip mechatronic microsystem for surface imaging and force response studies*
S. Hafizovic, D. Barrettino, T. Volden, J. Sedivy, K.-U. Kirstein, O. Brand and A. Hierlemann, *Proc Natl Acad Sci USA*. 2004, 101(49), pp. 17011-17015
 10. *Process simulation system for 3D X-ray lithography and development*
Y. Hirai, **S. Hafizovic**, J.G. Korvink and O. Tabata, *Transactions of the Institute of Electrical Engineers of Japan,-Part-E*. Sept. 2003, 123-E(9), pp. 368-375
 11. *Temperature-dependent thermal conductivities of CMOS layers by micro-machined thermal van der Pauw test structures*
S. Hafizovic and O. Paul, *Sensors and Actuators*, A97-98, 2002, pp. 246-252

Book Chapters

1. *Integrated cantilevers and atomic force microscopes*
S. Hafizovic, K.-U. Kirstein and Andreas Hierlemann, *Applied Scanning Probe Methods V*, Editor B. Bhushan, pp. 1-22, Springer (2007)
2. *Exposure simulation for LIGA*
J.G. Korvink, **S. Hafizovic**, P. Meyer, Y. Hirai, *Advanced Micro and Nanosystems*, Editors O. Brand, G. K. Fedder, C. Hierold, J.G. Korvink, O. Tabata, Wiley-VCH, expected 2007

Conference Contributions

1. *An 11k-electrode 126-channel highdensity microelectrode array to interact with electrogenic cells*
U. Frey, F. Heer, R. Pedron, **S. Hafizovic**, F. Greve, J. Sedivy, K.-U. Kirstein and A. Hierlemann, *ISSCC 2007*, San Francisco, USA, pp. 158 - 159
2. *11,000 electrode-, 126 channel-CMOS microelectrode array for electrogenic cells*

-
- U. Frey, F. Heer, R. Pedron, F. Greve, **S. Hafizovic**, K.-U. Kirstein and A. Hierlemann, MEMS 2007, Kobe, Japan, pp. 541 - 544
3. *Signal processing chain for CMOS microelectrode arrays to interface electrogenic cells*
S. Hafizovic, F. Heer and A. Hierlemann, Proc MEA Meeting, Reutlingen 2006
 4. *Integrated bidirectional 128-electrode array for neuronal cells*
S. Hafizovic, F. Heer, T. Ugniwenko, A. Blau and A. Hierlemann, FENS, Vienna, 2006
 5. *Prototype of a novel multielectrode recording and stimulation CMOS device for studying information processing in natural networks in vitro*
A. Blau, H. Baltes, P. Bonifazi, K. Brander, O. Dubochet, W. Franks, F. Greve, **S. Hafizovic**, F. Heer, T. Hessler, A. Hierlemann, B. Kochte-Clemens, S. Koster, T. Kraus, C. Leister, E. Ruaro, N. De Rooij, V. Torre, E. Verpoorte and C. Ziegler, FENS, Vienna 2006
 6. *CMOS bidirectional electrode array for electrogenic cells*
S. Hafizovic, F. Heer, W. Franks, F. Greve, A. Blau, C. Ziegler and A. Hierlemann, Proc. IEEE-MEMS 2006, Istanbul, Turkey, pp. 4-7
 7. *Single-chip atomic force microscope*
S. Hafizovic, T. Volden, D. Barrettino, K.-U. Kirstein and A. Hierlemann, Proc. IEEE-MEMS 2005, Miami, USA, pp. 638-641
 8. *CMOS monolithic atomic force microscope*
D. Barrettino, **S. Hafizovic**, T. Volden, J. Sedivy, K. Kirstein, A. Hierlemann, and H. Baltes, Proc. Symposium on VLSI Circuits., 2004, pp. 306-309
 9. *A single-chip CMOS micro-hotplate array for hazardous-gas detection and material characterization*
D. Barrettino, M. Graf, **S. Hafizovic**, S. Taschini, C. Hagleitner, A. Hierlemann and H. Baltes, Proc. ISSCC IEEE, 2004, Vol 1, pp. 312-313
 10. *A new platform for studying information processing by natural neural networks*
H. Baltes, A. Blau, P. Bonifazi, K. Brander, O. Dubochet, W. Franks, F. Greve, F. Heer, **S. Hafizovic**, T. Hessler, A. Hierlemann, B. Kochte-Clemens, S. Koster, T. Kraus, C. Leister, N.F. de Rooij, E. Ruaro, V. Torre, C. Ziegler, Proc. MEA Meeting, Reutlingen, 2004, p. 20
 11. *CMOS microchip for recording and stimulation of electrogenic cells*
F. Heer, W. Franks, A. Blau, **S. Hafizovic**, F. Greve, S. Taschini, C. Ziegler, A. Hierlemann and H. Baltes, Proc MEA Meeting, Reutlingen, 2004, p. 60.
-

12. *Effect of development process in moving mask deep X-ray lithography for 3D microfabrication*
N. Matsuzuka, Y. Hirai, O. Tabata, **S. Hafizovic** and J.G. Korvink, Proc. HARMST, 2003
13. *Measurement of PMMA dissolution rate and system calibration for predictive 3-D simulation of moving mask deep X-ray lithography*
Y. Hirai, **S. Hafizovic**, O. Tabata and J.G. Korvink, Proc. HARMST, 2003 Paper
14. *X3D: 3D X-ray lithography and development simulation for MEMS*
S. Hafizovic, Y. Hirai, O. Tabata and J.G. Korvink, Proc. Transducers'03, 2003, Vol 2, pp. 1570-1573
15. *3D simulation system for moving mask deep X-ray lithography*
Y. Hirai, N. Matsuzuka, **S. Hafizovic**, J.G. Korvink and O. Tabata, Proc. Micromechatronics and Human Science (MHS2003), 2003, pp. 271-276
16. *Temperature-dependent thermal conductivities of CMOS layers by micro-machined thermal van der Pauw test structures*
S. Hafizovic and O. Paul, Proc. Transducers '01, Vol 2, 2001, pp. 1398-401
17. *SPICE simulation of a thermal thin pattern recognition array*
R. Osorio, H. Ernst, M. Setiabudi, **S. Hafizovic**, G.A. Urban and J.G. Korvink, Proc. Eurosensors XIII, The Hague, NL, 1999, pp. 925-928
18. *Parsing for data exchange in coupled MEMS CAD*
H. Pham, **S. Hafizovic** and J. G. Korvink, Proc. IEEE CICC '99, Greece, 1999, pp. 258-263

Appendix C

Acknowledgements

First, I would like to thank my supervisors Andreas Hierlemann and Henry Baltes. Henry Baltes is the one who, in the first place, told me about neurons & chips while having o-toro & rice in Japan. I thank Andreas Hierlemann and Henry Baltes for having had the opportunity of pursuing my PhD at the Physical Electronics Laboratory. I joined a team-oriented and professionally run lab sparkling with bold ideas backed by competence. The lab closest to the PhD's dream environment that I have yet seen.

I am very much indebted to the dissertation's co-examiner Jan G. Korvink for countless discussions, inputs and motivators along the way to the PhD - both with regard to the dissertation and the broader aspects of life. I thank my co-examiner Sebastian Seung for ingenious input along the dissertation, especially during his sabbatical in Zurich.

Special thanks go to my outstanding office mates and friends Flavio Heer and Urs Frey. Flavio was already deep in the neurochip project when I came to join and he was my guide to the field. The day Urs joined in, the group took a leap in sophistication and professionalism. With and through these two remarkable swiss men, I lived four fantastic PhD years - be it in HPT, on foot, on two wheels, on two planks, in Chants, on Piz Kesch, or Prigel.

Thanks also go to my other project collaborators Wendy Franks and Frauke Greve on the neurochip, Diego Barrettino and Tormod Volden on the single-chip AFM and Stefano Taschini on digital design. Both projects were beyond of what can be obtained in one PhD's time and could only be realized by teamwork among several PhD students. Wendy Franks figured out how to get those platinum electrodes onto a chip in a way that even neurons were happy, Frauke Greve never tired of starting yet another culture and always knew how to tweak our lithography. Diego Barrettino would with a reach into his rich repository of circuits provide whatever amp or filter is required, just as Tormod Volden would carve AFM cantilevers from a Si crystal and magically mount tips so small that the naked eye cannot distinguish them.

A number of students are acknowledged for contributions to the dissertation: Dominic Frutiger and Stefan Schild for spike sorting, Jan Šedivý for creating the PCBs for the single-chip AFM, Agabi E. Oshioyoye for wafer processing and exorbitant Gedankenexperimente, David Moser for coding on Neurotalker, Adrian Bürli for AFM simulations and working on a resonant-mode setup for the single-chip AFM, Beat Hofstetter for coding on Neurotalker and for LSM simulations, Patrick Nüesch for coding spike correlation analysis in Neurotalker, Peter Friedli and Roger Ulrich for coding on Neurotalker.

Carlota Diaz is acknowledged for help with confocal scanning microscopy on our funny neurochips. The Kaiserslautern project collaborators group headed by Christiane Ziegler and the die-hard team consisting of Axel Blau and Tanja Ugniwenko, who never tired in trying different cultures on our neurochips until we finally found the ultimate recipe for creating little neuro computers.

Not much would happen without Erna Hug and Yellena von Almen in the administration, Igor Levak in the computer system and Donat Scheiwiler handling the laboratory infrastructure. Thanks to all the other PELers I shared time with: Oliver Brand, Nils Gödecke, Nicole Kerness, Kay-Uwe Kirstein, Adrian Kümmerli, Petra Kurzawski, Yue Lee, Jan Lichtenberg, Markus Graf, Christoph Hagleitner, Branka Rošćić, Tomi Salo, Jürg Schwizer, Wan Ho Song, Robert Sumier, Cyril Vancura, Tobias Vancura, and Martin Zimmermann.

Outside the lab environment, there are many people as crucial, a few of which I want to thank here. Thanks to my flat mates for deliberate and needed distraction, Jörg Rychen for countless good times and many thriving discussions ranging from fish platters to quantum dots to 3D calculation machines, Daniel Müller for spiritual company and for reminding me of the world outside Zurich. Finally, I am vastly indebted to my dear family for fathomless backing.

Curriculum Vitae

Sadik Hafizović

Born March 16, 1977, Berlin, Germany

German Citizen

Sept. 2006 Defense of the dissertation entitled *Neural Interface and Atomic-force Microscope in CMOS Technology*.

2002 - 2006 Work on Ph.D. thesis and related topics at the Physical Electronics Laboratory, ETH Zurich, under the supervision of Prof. Baltes and Prof. Hierlemann.

2001 - 2002 Diploma thesis: Ritsumeikan University Kyoto, Japan.

1997 - 2002 Dipl.-Ing. Microsystems, IMTEK, Freiburg, Germany .

1996 Abitur.

1994 - 1996 Gymnasium Marktheidenfeld, Germany.

1993 - 1994 Gymnasium Bad Gandersheim, Germany.

1992 - 1993 Saudi-American International High School Al Hada, Saudi Arabia.

1987 - 1992 Gymnasium Bad Gandersheim, Germany.

1984 - 1987 Primary School Bad Gandersheim, Germany.